Biolinguistic Explorations: Design, Development, Evolution

Noam Chomsky

Abstract

Biolinguistic inquiry investigates the human language faculty as an internal biological property. This article traces the development of biolinguistics from its early philosophical origins through its reformulation during the cognitive revolution of the 1950s and outlines my views on where the biolinguistic enterprise stands today. The growth of language in the individual, it is suggested, depends on (i) genetic factors, (ii) experience, and (iii) principles that are not specific to the faculty of language. The best current explanation of how language is recursively generated is through Merge, an operation that takes objects already constructed, and reconstructs a new object from them, generating a ‘language of thought’, perhaps in a manner close to optimal (relying on principles of category (iii)), with externalization (hence communication) a secondary process. The concluding section of the article offers several objectives for future research in the field.

Keywords: biolinguistics; evolution of language; I-language; language of thought; semantic interface; Merge

The problem that has virtually defined the serious study of language since its ancient origins, if only implicitly, is how to identify the specific nature of this distinctive human possession. Within the ‘biolinguistic perspective’ that began to take shape fifty years ago, the concern is transmuted into the effort to determine the genetic endowment of the faculty of language, understood to be a ‘cognitive organ’, in this case virtually shared among humans and in crucial respects unique to them, and hence a kind of species property. So construed, language is I-language (internal language), a state of the computational system of the mind/brain that generates structured expressions, each of which can be taken to be a set of instructions for the interface systems within which the faculty of language is embedded. There are at least two such interfaces: the systems of thought that use linguistic expressions for reasoning, interpretation, organizing action, and other mental acts; and the sensorimotor systems that externalize expressions in production and construct them from sensory data in perception. The theory of the genetic
endowment for language is commonly called ‘Universal Grammar’, UG, to adapt traditional terms to a different framework. Certain configurations are possible human languages, others are not, and a primary concern of the theory of human language is to establish the distinction between the two categories. The term ‘biolinguistics’ itself was coined in 1974 as the topic for an international conference that brought together evolutionary biologists, neuroscientists, linguists, and others concerned with language and biology, one of many such initiatives, before and since.

Within the biolinguistic framework, several tasks immediately arise. The first is to construct generative grammars for particular languages that yield the facts about sound and meaning, the correct ‘instructions’ for the interface systems; what has been called the problem of ‘descriptive adequacy’. It was quickly learned that the task is formidable. Very little was known about languages, despite millennia of inquiry. The most extensive existing grammars and dictionaries were, basically, lists of examples and exceptions, with some weak generalizations. When the issue was even addressed, it was assumed that anything beyond could be determined by unspecified methods of ‘analogy’ or ‘induction’. But even the earliest efforts revealed that these notions concealed vast obscurity. Traditional grammars and dictionaries tacitly appeal to the understanding of the reader, either knowledge of the language in question or the shared innate linguistic capacity, or more commonly both. But for the study of language as part of biology, it is precisely that presupposed understanding that is the topic of investigation, and as soon as the issue was faced, major problems were quickly unearthed.

The second task is to account for the acquisition of language. In biolinguistic terms, that means discovering the operations that map presented data to the I-language attained. Abstractly formulated, it is the problem of constructing a ‘language acquisition device’ (LAD), the problem of ‘explanatory adequacy’. With sufficient progress in approaching explanatory adequacy, a further and deeper task comes to the fore: to transcend explanatory adequacy, asking not just what the mapping principles are, but why language growth is determined by these principles rather than innumerable others that can easily be imagined. The question was premature until quite recently, when it has been addressed in what has come to be called the minimalist programme, the natural next stage of biolinguistic inquiry.

The questions arise in a natural succession in the course of research: in brief, the ‘what’ questions (descriptive adequacy), the ‘how’ questions (explanatory adequacy), and the ‘why’ questions (beyond explanatory adequacy). Of course, there is constant interaction. Even speculative answers to the later questions inform the quest for answers to the earlier ones.

Another question is how the faculty of language evolved. There are libraries of books and articles about the evolution of language – in rather striking contrast to the literature, say, on the evolution of the communication systems of bees. For human language, the problem is vastly more
difficult for obvious reasons, and can be undertaken seriously, by definition, only to the extent that some relatively firm conception of UG is available – that is, a conception of what has evolved.

Many other questions arise: about how language is put to use, the neural structures implicated in knowledge and use of language, and much more.

I would like to say a few words on where I think the biolinguistic enterprise stands today, keeping mostly to some leading ideas, some of ancient vintage, which tend to be obscured in details of implementation, and also straying a little into areas that are not narrowly linguistic, but that have for centuries been considered closely related to language, with an interesting revival in recent years, particularly in the domain of moral philosophy and its significance for human affairs.

Inquiry into these two domains proceeds along almost parallel paths: almost parallel because they should meet short of infinity, though where is far from clear. One path seeks to understand more about language and mind. The other is guided by concerns for freedom and justice. There should be some shared elements, in particular, what the co-founder of modern evolutionary theory, Alfred Russel Wallace, called ‘man’s intellectual and moral nature’: the human capacities for creative imagination, language and symbolism generally, interpretation and recording of natural phenomena, intricate social practices and the like, a complex of capacities that seem to have crystallized fairly recently among a small group in East Africa of which we are all descendants, sometimes called simply ‘the human capacity’. The archaeological record suggests that the crystallization was sudden in evolutionary time. Some eminent scientists call the events ‘the great leap forward’, which distinguished contemporary humans sharply from other animals. It is commonly assumed that the emergence of language was a core element in the great leap forward, and language is one component of the human capacity that is accessible to study in some depth. The principles of our intellectual and moral nature remain a considerable mystery, but we can hardly doubt that they are rooted in our very nature and play a central role in our lives. This is conventionally denied, but not credibly, in my opinion.

Puzzlement about the human capacity goes back to the origins of reflective thought in every culture we know much about. In the West, there was another kind of ‘great leap forward’ in the course of the scientific revolutions of the seventeenth century, which set the stage for modern science, and also included a cognitive revolution that is a precursor of what is commonly called ‘the cognitive revolution of the 1950s’, better considered a second cognitive revolution, which rediscovered themes that had been developed in thoughtful and instructive ways from which we still have a lot to learn.

One theme, developed by the physician/philosopher Juan Huarte in the sixteenth century, is that the essential property of human intelligence is the capacity of the human mind to ‘engender within itself, by its own power,
the principles on which knowledge rests’ (Huarte, 1575). Recognition that the capacity to gain understanding relies on fixed principles of the mind is as old as Plato, who could suggest only that the principles are remembered from an earlier existence. The seventeenth-century cognitive revolution sought to purge Plato’s theory of ‘the error of reminiscence’, as Leibniz put it, which means attributing to the mind itself an invariant innate structure. We would reformulate that today in terms of some genetic endowment, which enables an animal to construct experience from chaotic stimuli, constructing what ethologists called an *Umwelt*, a world of experience, differing for us and bees; and to rely on internal principles to construct our conception of the world in which we live, including language, moral codes, and a great deal more.

For Cartesian scientists, as in fact for Huarte much earlier, the fundamental nature of human intelligence was a kind of creative capacity, exhibited most clearly in the use of language to express thought, a basic distinction between man and beast for the Cartesians: the capacity of every person, but no beast or machine, to construct linguistic expressions without bound, free of external stimuli or internal physical states, appropriate to circumstances but not caused by them – a fundamental distinction – and eliciting thoughts in others that they could have expressed the same way. Cartesian psychologists developed experimental tests to determine whether another creature has these creative capacities, and concluded that if it passes the strictest tests we can devise, it would only be reasonable to conclude that it has a mind like ours: their proposed solution to ‘the problem of other minds’. Notice that this is a more sophisticated and scientifically sound version of what has come to be called ‘the Turing test’ in recent years, based on a serious misunderstanding of a brief paper of Turing’s, I believe (Turing, 1950). Such ideas contributed directly to investigation of the nature of concepts, some of them ‘common notions’ that are part of our nature, developed or refined through experience. They also contributed to investigation of the principles by which linguistic expressions are constructed, the research programme of ‘philosophical grammar’, meaning what we would call scientific grammar, keeping to the science–philosophy distinction established in the nineteenth century.

Hume recognized that knowledge and belief are grounded in a ‘species of natural instincts’, part of the ‘springs and origins’ of our inherent mental nature. He also recognized that something similar must be true in the domain of moral judgment. Our moral judgments are unbounded in scope and we constantly apply them in systematic ways to new circumstances. Hence they too must be founded on general principles that are part of our nature though beyond our ‘original instincts’, more primitive instincts, perhaps shared with animals. That should lead directly to efforts to develop something like a grammar of moral judgment. That task was undertaken by the leading American moral and political philosopher of the late twentieth
century, John Rawls, who used the analogy of linguistic models of generative grammar being developed in the 1960s. In the last few years, that has become a lively field of theoretical and empirical inquiry.

It is worth clarifying that the questions can be sensibly raised in this form only if we regard the human capacity as an internal property of people – part of their organic nature, on a par with the digestive and immune systems, the systems of visual interpretation and motor organization, and so on – as ‘an organ of the body’, in the loose informal sense in which the term ‘organ’ is conventionally used: some subcomponent of a complex organic system that has sufficient internal integrity so that it makes sense to study it in abstraction from its complex interactions with other systems in the life of the organism. The biolinguistic perspective developed within this framework, taking the faculty of language to be a ‘mental organ’, where the term ‘mental’ simply refers to certain aspects of the world, to be studied in the same way as chemical, optical, electrical, and other aspects.

For Cartesian scientists, the domain of the mental was distinct from body, the physical world. That was a reasonable scientific thesis based on the plausible belief that elements of the human capacity, specifically the creative aspect of language use, could not be explained in terms of physical mechanisms: the ‘mechanical philosophy’, in the terminology of the day. That belief was undermined by Newton, who demonstrated, to the consternation of leading scientists including Newton himself, that we have no coherent notion of machine, body, physical, material, beyond the best theories of various aspects of the world that we can devise, always looking forward to deeper explanation and eventual unification, though in unpredictable and what have turned out often to be quite surprising ways. As historians of science have recognized, this intellectual move ‘set forth a new view of science’ in which the goal is ‘not to seek ultimate explanations’ but to find the best theoretical account we can of the phenomena of experience and experiment (I. Bernard Cohen). It follows that we have no choice but to adopt some non-theological version of what historians of philosophy call ‘Locke’s suggestion’: that God might have chosen to ‘superadd to matter a faculty of thinking’ just as he ‘annexed effects to motion which we can in no way conceive motion able to produce’ (Locke, 1689/1975: p. 541) – notably the property of action at a distance, a revival of occult properties, many leading scientists argued (with Newton’s partial agreement). The mind–body problem in its traditional sense therefore became unformulable, because the notion of body disappeared.

All of this was pretty well understood through the eighteenth and nineteenth centuries, leading to the conclusion that properties ‘termed mental’ reduce somehow to ‘the organical structure of the brain’, in the formulation of chemist/philosopher Joseph Priestley in the late eighteenth century. However, it seems to have been forgotten. The American Academy of Arts and Sciences published a volume summarizing the results of the Decade of
the Brain that ended the twentieth century. The guiding theme, formulated by the distinguished neuroscientist Vernon Mountcastle, is the thesis of the new biology that ‘Things mental, indeed minds, are emergent properties of brains, [though] these emergences are ... produced by principles that ... we do not yet understand’ (Mountcastle, 1998: p. 136). The same thesis has been put forth in recent years by prominent scientists and philosophers as an ‘astonishing hypothesis’ of the new biology, a ‘radical’ new idea in the philosophy of mind, ‘the bold assertion that mental phenomena are entirely natural and caused by the neurophysiological activities of the brain’, opening the door to novel and promising inquiries, a rejection of Cartesian mind–body dualism, and so on. All reiterate formulations of centuries ago, in virtually the same words, after mind–body dualism became unformulable with the disappearance of the only coherent notion of body (physical, material, etc.) – facts well understood in standard scholarly work, for example, Friedrich Lange’s classic nineteenth-century history of materialism.

Although the traditional mind–body problem dissolved after Newton, the phrase ‘mind–body problem’ has been resurrected for a problem that is only loosely related to the traditional one. The traditional mind–body problem developed in large part within normal science: certain phenomena could not be explained by the principles of the mechanical philosophy, the presupposed scientific theory of nature, so a new principle was proposed, some kind of res cogitans alongside of material substance. The next task would be to discover its properties and to try to unify the two substances. That task was undertaken, but was effectively terminated when Newton undermined the notion of material substance.

What is now called ‘the mind–body problem’ is quite different. It is not part of normal science. The new version is based on the distinction between the first-person and the third-person perspective. The first-person perspective yields a view of the world presented by one’s own experience – what the world looks like, feels like, sounds like to me, and so on. The third-person perspective is the picture developed in its most systematic form in scientific inquiry, which seeks to understand the world from outside any particular personal perspective.

The new version of the mind–body problem resurrects some observations of Bertrand Russell’s eighty years ago, though the concerns have a long history, going back to classical Greece. Democritus presents an argument between the intellect and the senses about what is real, where the former says ‘Ostensibly there is color, ostensibly sweetness, ostensibly bitterness, actually only atoms and the void’, and the senses reply, ‘Poor intellect, do you hope to defeat us while from us you borrow your evidence? Your victory is your defeat’ (cited in Schrödinger, 1958: p. 163). Exploring these ideas, Russell asked us to consider a blind physicist who knows all of physics but doesn’t know something we know: what it’s like to see the colour blue. Russell’s conclusion was that the natural sciences seek to discover ‘the
causal skeleton of the world’ (Russell, 1927: pp. 391–2). Other aspects lie beyond their purview.

Recasting Russell’s thought experiment in naturalistic terms, we might say that like all animals, we have internal capacities that reflexively provide us with an *Umwelt*, a world of experience. But being reflective creatures, thanks to the emergence of the human capacity, we go on to seek to gain a deeper understanding of the phenomena of experience. These exercises are called myth, or magic, or philosophy, or science, in its modern sense. If humans are part of the organic world, we can expect that our capacities of understanding and explanation will have fixed scope and limits – a truism that is sometimes thoughtlessly derided as ‘mysterianism’, though it was recognized by Descartes and Hume, among others. It could be that these innate capacities do not lead us beyond some understanding of Russell’s causal skeleton of the world, and it is always an open question how much of that can be attained. In principle, these could become topics of empirical inquiry into the nature of what we might call ‘the science-forming faculty’, another ‘mental organ’. These are interesting topics, but the issues are distinct from traditional dualism, which evaporated after Newton.

It might also be noted that after the disappearance of traditional metaphysical dualism, a new form of ‘methodological dualism’ has developed, similar to metaphysical dualism in its scope. But unlike the scientific doctrine of metaphysical dualism, methodological dualism is a pernicious dogma, which takes many forms, and has, I believe, caused great harm in contemporary philosophy, matters I discussed in this publication some years ago (Chomsky, 1994).

The cognitive revolution of the seventeenth century also led to inquiry into the nature of concepts, with important contemporary implications, also insufficiently appreciated. Aristotle had recognized that the objects to which we refer in using language cannot be identified by their material substance. A house, he pointed out, is not merely a collection of bricks and wood, but is defined in part by its function and design: a place for people to live and store their possessions, and so on. In Aristotle’s terms, a house is a combination of matter and form. Notice that his account is metaphysical: he is defining what a house *is*, not the word or idea ‘house’. That approach led to hopeless conundrums. The ship of Theseus is a classic case; Saul Kripke’s puzzle about belief is a contemporary analogue (Kripke, 1979). With the cognitive turn of the seventeenth century these questions were reframed in terms of operations of the mind: what does the word ‘house’ mean, and how do we use it to refer? Pursuing that course we find that for natural language there apparently is no word–object relation, where objects are mind-independent entities. That becomes very clear for Aristotle’s example, the word ‘house’, when we look into its meaning more closely. Its ‘form’ in the Aristotelian sense is vastly more intricate than he assumed.
One may be misled by the fact that houses are artifacts, so let’s consider simpler cases that were investigated in seventeenth- and eighteenth-century classics. Take ‘river’, a notion that Hobbes considered. He suggested that rivers are mentally individuated by origin. But while there is some truth to the observation, it is not really accurate, and it only scratches the surface of our intuitive understanding of the concept. Consider the Charles River, which flows near my office. It would remain the very same river under quite extreme changes – among many others, reversing its course – so origin cannot be the defining characteristic. It would still be the Charles River, under some circumstances, if it were divided into separate streams that converged in some new place, or if any \( \text{H}_2\text{O} \) that happens to be in it were replaced by chemicals from an upstream manufacturing plant. On the other hand, under trivial changes it would no longer be a river at all: for example, directing it between fixed boundaries and using it for shipping freight (in which case it would be a canal, not a river), or hardening the surface to the glassy state by some near-undetectable physical change, painting a line down the middle, and using it to drive to Boston (in which case it would be a highway). And on to further intricacies, as we can easily determine.

Much the same holds of even the simplest concepts – lexical items, in linguistic terms. Exploring them confirms the conclusion that the world of experience and interpretation is constructed by our rich ‘cognoscitive powers’, as they were called, relying on gestalt properties, causal relations, and other components of mental operations that were explored by seventeenth-century British philosophers and their successors. Summarizing many years of discussion of such topics, Hume concluded that ‘the identity we ascribe’ to minds, vegetables, animal bodies, and other entities is ‘only a fictitious one’ established by the imagination, not a material nature belonging to this form, a conclusion that appears to be basically correct (Hume, 1739).

With a shift from a metaphysical to a cognitive interpretation of the nature and use of referring expressions, familiar conundrums dissolve. There is no reason to expect our cognoscitive powers to have an answer to every question that can be raised in any situation that can be contrived: for example, the question whether the ship of Theseus is what is built on shore from the original boards that were cast successively to sea, or what is sailing with the boards having been successively replaced. And there is no reason why London cannot be pretty from one perspective and not pretty from another, even if two names for London are used, since there is no mind-independent entity \textit{London} – which seems clear enough on other grounds: what object of the world could be simultaneously both concrete (it can be destroyed by an earthquake) and abstract (it can be rebuilt somewhere else, with different materials, maybe after hundreds of years), merely to scratch the surface of the mental construction ‘London’? The same holds of other cases; probably all, if we keep to natural language, instantiations of UG,
sorting it out from the heterogeneous complexes called ‘language’ in informal usage.

A notion of particular significance to Locke and his successors was ‘person’. Locke recognized that persons are individuated in part by psychic continuity. Thus a single body could house two distinct persons if two distinct psychic continuities were connected to it, and a person could persist even if its substance became completely different, properties of our concepts exploited by science fiction. These are all properties of the common notion or innate idea of person. Locke argued further that ‘person’, unlike ‘tree’ or ‘cat’, is a ‘forensic term, appropriating actions and their merit’, belonging only to ‘intelligent agents, capable of a law, and happiness, and misery’, and hence even more remote from mind-independent physical investigation, and at the heart of our moral faculty and intuitions (Locke, 1689/1975: p. 346).

These properties of concepts can hardly be acquired by experience, no matter how rich. And we would expect them to be shared in essentials among languages and cultures, though some variation is tolerated; semantic fields, such as colour or notions related to thought (‘know’, ‘believe’, and the like) can divide the domain in somewhat different ways. Furthermore, these properties are available to very young children. A lot of children’s literature is based on them; in the case of psychic continuity, the prince turned by an evil witch into a frog, who comes to look like the prince again when he is kissed by the beautiful princess, though he was the prince all along. But Locke’s observations are too narrow. The basic principles extend well beyond ‘person’. My grandchildren are entranced by the story of a baby donkey, named Sylvester, who is somehow turned into a rock and spends the rest of the story trying to convince his parents that he is really their child. Since children’s stories have happy endings, at the end he is turned back into the baby donkey, and everyone lives happily ever after. But every child understands that the rock was really Sylvester all along, no matter what its material constitution was.

In all such cases, there is no mind-independent object, which could in principle be identified by a natural scientist, related to the name; or to other words and phrases. As we proceed, we find much more intricate properties, no matter how simple the terms of language we investigate. For natural language and thought there appears to be no meaningful word–object relation because we do not talk about the world in terms of mind-independent objects; rather, we focus attention on intricate aspects of the world by resort to our cognoscitive powers. If so, then for natural language and thought there is no notion of reference in the technical sense of Frege, Peirce, Tarski, Carnap, and others, or contemporary philosophical externalists such as Putnam and Kripke. That fact does not inhibit the act of referring: I can tell you that London was bombed in July, using the internal resources of my mind to direct your attention to particular aspects of the
world, communicating successfully insofar as our cognoscitive powers and relevant experience are shared; or I can refer successfully to London by just lamenting an awful tragedy, never using any phrase that has a special relation to London. Those are normal acts of referring, not requiring some relation with the properties of the technical notions reference or denotation. These technical concepts are fine for the context for which they were invented: formal systems where the symbols, objects, and relations are stipulated. Arguably they also provide a norm for science: its goal is to construct systems in which terms really do pick out an identifiable mind-independent element of the world, like ‘neutron’ or ‘noun phrase’. But human language and common-sense thought do not work that way. The situation is complex with regard to the natural sciences too. Take the concepts ‘individual’ and ‘phenotype’, central to biology. From other no less scientific perspectives, there is no particular significance to what is happening within the skin, a curious construction whose material components are changing every instant. But the concepts are not problematic in the particular natural-science contexts in which they are used.

In these respects, the internal conceptual symbols of human language and thought are rather like the phonetic units of mental representations, such as the syllable [ba]. Every particular act externalizing this internal mental object yields a mind-independent entity, but it is idle to seek a mind-independent construct that corresponds to the syllable. Communication is not a matter of producing some mind-external entity that the hearer picks out of the world, the way a natural scientist could. Rather, communication is a more-or-less affair, in which the speaker produces external events and hearers seek to match them as best they can to their own internal resources. Words and concepts appear to be similar in this regard, even the simplest of them. Communication relies on largely shared cognoscitive powers, and succeeds insofar as similar mental constructs, background, concerns, and presuppositions allow for similar perspectives to be reached. If that is true – and the evidence seems overwhelming – then natural language diverges sharply in these elementary respects from animal communication, which appears to rely on a one–one relation between mind/brain processes and ‘an aspect of the environment to which these processes adapt the animal’s behavior’, to quote cognitive neuroscientist C. R. Gallistel’s introduction to a volume of papers on animal communication (Gallistel, 1990: p. 15).

If this picture is essentially accurate, then one basic problem facing the study of evolution of the human capacity is to account for the emergence of these universal and innate properties of language and thought; not a trivial matter, as we quickly discover when we take words and ideas seriously.

When we move beyond the simplest elements of language, problems of this kind escalate rapidly. That was discovered as soon as the long-forgotten tradition was revived in the 1950s. To give a few of the early illustrations for concreteness, the internal linguistic resources that most English speakers
more or less share determine that the sentence ‘Mary saw the man leaving
the store’ is three-ways ambiguous, though it may take a little reflection to
reveal the fact. But the ambiguities are resolved if we ask ‘Which store did
Mary see the man leaving?’, understood approximately as ‘Which store did
Mary see the man leave?’ The phrase ‘which store’ is raised from the posi-
tion in which its semantic role is determined as object of ‘leave’, and is then
given an additional interpretation as an operator taking scope over a vari-
able in its original position, so the sentence means, roughly, ‘for which x, x
a store, Mary saw the man leave(ing) the store x’ – and without going into it
here, there’s good reason to suppose that the semantic interface really does
interpret the variable x as ‘the store x’, a well-studied phenomenon called
‘reconstruction’. The phrase that serves as the restricted variable is silent in
the phonetic output, but must be there for interpretation. Only one of the
underlying structures permits the operation, so the ambiguity is resolved in
the interrogative, in the manner indicated. The constraints involved – so-
called ‘island conditions’ – have been studied intensively for about forty-five
years. Recent work indicates that they may reduce in large measure to
minimal search conditions of optimal computation, perhaps not coded in
UG but more general laws of nature – which, if true, would carry us beyond
explanatory adequacy to the domain of ‘why’ questions.

To take a second example, consider the sentence ‘John ate an apple.’ We
can omit ‘an apple’, which yields ‘John ate’, which we understand to mean
‘John ate something unspecified.’ Now consider ‘John is too angry to eat an
apple.’ We can omit ‘an apple’, which yields ‘John is too angry to eat’, which,
by analogy to ‘John ate’, would be expected to mean that John is so angry
that he won’t eat anything. That is in fact a natural interpretation, but there
is also a second one in this case: roughly, John is so angry that someone or
other won’t eat him, John – the natural interpretation for the structurally
analogous expression ‘John is too angry to invite.’ In this case, the explana-
tion appears to lie in the fact that the phrase ‘too angry to eat’ does include
the object of ‘eat’, but it is invisible. The invisible object is raised just as
‘which store’ is raised in the previous example, again yielding an operator–
variable structure. In this case, however, the operator has no content, so the
construction is an open sentence with a free variable, and hence a predicate.
The semantic interpretation follows from general principles. The minimal
search conditions that restrict the raising of ‘which store’ in the first example
also bar the raising of the empty object of ‘eat’, which yields standard island
properties. Again, there is substantial independent evidence supporting
such conclusions, for a variety of constructions.

In both cases, then, general computational principles yield the required
interpretations as an operator–variable construction, with the variable
unpronounced in both cases and the operator unpronounced in one. The
surface forms in themselves tell us little about the interpretations. That is a
common situation. For such reasons, it was recognized in the earliest work
in generative grammar that the language that every person quickly masters relies on inner resources to generate internal expressions that yield information of the kind just illustrated. Even the simplest elements that enter into sound and meaning – phonemes and morphemes, in technical terms – find their place in the generative procedures that yield the expressions, but cannot in general be detected in the physical signal; they may even be silent, as in the syntactic/semantic examples just given. For that reason, it seemed – and seems – that the language acquired must have the basic properties of an internalized explanatory theory. These are design properties that must be dealt with by an account of the growth of language in the individual – language acquisition – and its evolution in the species, alongside of the elementary semantic properties of the simplest words.

Quite generally, construction of theories must be guided by what Charles Sanders Peirce a century ago called an abductive principle, which he took to be a genetically determined instinct, like the pecking of a chicken. The principle ‘puts a limit upon admissible hypotheses’, so that the mind is capable of ‘imagining correct theories of some kind’ and discarding infinitely many others consistent with available evidence (Peirce, 1957: p. 238). Peirce was concerned with what I was calling ‘the science-forming faculty’, but similar problems arise for language acquisition, though it is dramatically unlike scientific discovery. It is rapid, virtually reflexive, convergent among individuals, relying not on controlled experiment or instruction but only on the blooming buzzing confusion that each infant confronts. The format that limits admissible hypotheses about structure, generation, sound, and meaning must therefore be highly restrictive. That much was recognized from the earliest modern work. It seems to follow that the language faculty must be richly articulated and specific to this cognitive system. Plainly such conclusions make it next to impossible to raise questions that go beyond explanatory adequacy – the ‘why’ questions – and pose serious barriers to inquiry into how the faculty might have evolved, the latter a problem discussed inconclusively at the 1974 conference.

In the 1950s, when this work began to take shape, the prevailing ideas in both linguistics and general biology adopted a particular interpretation of what may be the most quoted passage in biology, Darwin’s concluding words of the first edition of *Origin of Species*: that ‘from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved’ (Darwin, 1859). A widely held view among evolutionary biologists was that the variability of organisms is so free as to constitute ‘a near infinitude of particulars which have to be sorted out case by case’ (Stent, 1985: p. 2). The counterpart in the study of language was the ‘Boasian’ conception that languages can ‘differ from each other without limit and in unpredictable ways’ (Joos, 1966: p. 228), so that the study of each language must be approached ‘without any preexistent scheme of what a language must be’ (Joos, 1966: p. 1). In neither case, of course, were the
formulaic statements meant literally (which would have been incoherent). But they did capture the spirit of prevailing conceptions, which had substantial motivation: languages and organisms do appear on the surface to have virtually endless variety.

To look with a broader sweep, the problem of reconciling unity and diversity has constantly arisen in biology and linguistics. The linguistics of the early scientific revolution distinguished universal from particular grammar, though not in the sense of the biolinguistic framework. Universal Grammar was taken to be the intellectual core of the discipline; particular grammars are accidental instantiations. With the flourishing of anthropological linguistics, the pendulum swung in the other direction, towards diversity, well captured in the ‘Boasian’ formulation. In general biology, a similar issue had been raised sharply in the Cuvier–Geoffroy debate in 1830. Cuvier’s position, emphasizing diversity, prevailed, particularly after the Darwinian revolution, leading to the conclusions about near infinitude of variety that have to be sorted out case by case. Over the years, in both general biology and linguistics, the pendulum has been swinging towards unity, with some interaction. There is reason to hope that the analogies may become more substantive in coming years. I will return to that.

The prevailing conceptions of half a century ago posed a kind of paradox, in both domains. For the study of language, considerations of acquisition appeared to show that the innate pre-existing scheme must be highly restricted so as to limit admissible hypotheses. But that runs counter to the apparently limitless variety of attainable languages. The paradox may be less glaring in general biology, but it is still there. In the study of language, about twenty-five years ago efforts to overcome the paradox converged in an approach that seemed to offer a promise of a conceptual breakthrough. This Principles and Parameters (P & P) approach proposed a sharp distinction between process of acquisition and the format of the internal theory of language. The basic idea is that the genetic endowment provides a fixed system of principles that interact in the generation of the expressions of a language, and a fixed set of parameters that are assigned a value by exposure to external data. A language is determined by setting the values and selecting lexical items from a fixed and highly structured store, with properties of the kind I have mentioned. The promissory note is to be paid by showing that the parameters are simple enough so that they can be set by the data available to children, and that the fixed principles, with parameters set, function automatically to yield the rich variety of expressions for every possible human language, with their specific interpretations. That’s a huge empirical task, of course, but one that has been pursued with some success, opening up a very wide range of typologically varied languages to much deeper study than before, identifying points of difference, revitalizing the study of language acquisition, exposing a host of new
problems for investigation, sometimes providing answers. As a research programme, at least, it has been quite successful.

The P & P approach largely emerged from intensive study of a range of languages, but it was also suggested by an analogy to new developments in biology, specifically the discovery of regulatory mechanisms by François Jacob and Jacques Monod, and Jacob’s speculations about how slight changes in these mechanisms might yield great superficial differences – a butterfly or an elephant, and so on. The model seemed natural for language as well: slight changes in parameter settings might yield superficial variety, through interaction of invariant principles with parameter choices (Chomsky, 1980: p. 67).

The approach also removed a crucial conceptual barrier to the study of evolution of language. In earlier approaches, it appeared that the genetically determined format for language had to be highly articulated, richly structured, and specific to language, sharply restricting admissible hypotheses so as to account for the rapidity and convergence of acquisition. Plainly, the greater the richness and specificity of the format, the harder it will be to account for its evolution. But by divorcing acquisition from the format for language, we overcome that conceptual problem. It may still turn out that the innate mechanisms are rich, highly structured, and specific to language, but that is no longer a conceptual necessity, within the P & P framework.

It therefore became possible to pursue more seriously the recognition, from the early days of generative grammar, that acquisition of language involves not just a few years of experience and millions of years of evolution, yielding the genetic endowment, but also ‘principles of neural organization that may be even more deeply grounded in physical law’ (Chomsky, 1965: p. 59). At the same time, what biologists call the evo-devo revolution was increasingly showing that ‘the rules controlling embryonic development’ interact with other physical conditions ‘to restrict possible changes of structures and functions’ in evolutionary development, providing ‘architectural constraints’ that ‘limit adaptive scope and channel evolutionary patterns’ (to quote Jacob and others). By the 1980s, evo-devo discoveries were reviving interest in ideas of rational morphology that predate Darwin, and observations in the early days of modern evolutionary biology that there appear to be ‘predetermined lines of modification’ that lead natural selection to ‘produce varieties of a limited number and kind’ for every species (Thomas Huxley), ideas that had been explored particularly by D’Arcy Thompson and Alan Turing. These were summarized in an influential ‘consensus paper’ by Maynard Smith and others twenty years ago, and have since been considerably extended by discoveries about organizing principles, deep homologies, and conservation of fundamental mechanisms of development, most famously hox genes. By now such considerations have been adduced for a wide range of problems of development and evolution, from cell division to optimization of structure and function of neural
networks, and just a few weeks ago, proposals as to how principles of physics determine that vertebrate brains have a certain distribution of gray and white matter, with specific characteristics. Such discoveries, once again, have suggestive similarities to ongoing linguistic investigation, in this case efforts to show that general principles of computational efficiency determine specific properties of the faculty of language, raising the question whether, or to what extent, the linguistic organ is optimally designed to satisfy conditions that it must satisfy to be usable at all.

Assuming that language has general properties of other biological systems, we should, therefore, be seeking three factors that enter into the growth of language in the individual:

1 Genetic factors, the topic of UG. These interpret part of the environment as linguistic experience, and determine the general course of development to the languages attained.
2 Experience, which permits variation within a fairly narrow range.
3 Principles not specific to the faculty of language.

The third factor includes principles of efficient computation, which would be expected to be of particular significance for systems such as language, determining the general character of attainable languages.

An elementary fact about the language faculty is that it is a system of discrete infinity: an I-language generates a discrete infinity of hierarchically structured expressions, submitted to interpretation. There are many ways to generate such discrete infinities. One way is Phrase Structure Grammar, which made sense when it was proposed because the various kinds of PSG captured some significant linguistic properties (like hierarchic structure and nested embedding), and also fit naturally into one of the standard versions of recursive function theory (Post’s theory). Over the years, it has been shown that the various intricacies of PSG are wrong or superfluous, and by now it seems to be possible to reduce the operations to the simplest possible form of recursive generation: an operation that takes objects already constructed, and constructs from them a new object. Call that operation Merge. If computation is efficient, then if X and Y are merged, neither will be changed; computation will observe the no-tampering condition. We can then take the result of Merge of X, Y to be simply the set \{X, Y\}. With Merge available, we instantly have an unbounded system of hierarchically structured expressions. As already noted, these expressions link to language-external but organism-internal systems, the systems of thought and the sensorimotor system. In traditional terms, a language pairs sound and meaning in a specific way. The expressions generated by the language have to satisfy these interface conditions. Insofar as third-factor properties function, language will satisfy these conditions in an optimal way, meeting conditions of efficient computation. We can regard an account of some
linguistic phenomena as *principled* insofar as it derives them by efficient computation satisfying interface conditions. A very strong assumption, sometimes called ‘the strong minimalist thesis’, is that all phenomena of language have a principled account in this sense, that language is a perfect solution to interface conditions. UG is the residue when this thesis is not satisfied. An account of the evolution of language will have to deal with the property of unbounded Merge, and whatever else remains in UG, the genetic endowment.

The operation Merge can be complicated by adding the principle of associativity, suppressing hierarchy, and if linear order is added, yielding strings of symbols. Those extra complications are in fact incorporated in the mapping to the sensorimotor system (SM), along with many others. However, for syntax-semantics those extra complications appear to be inappropriate. Expressions keep to the simplest generated form: hierarchically structured. It is in this form that they are submitted to semantic interpretation, as in the simple examples already discussed. Apart from externalization (the functioning of articulators, etc.), there is little interest in unstructured strings of symbols, the product of ‘weak generation’, in the terminology of mathematical linguistics. Or in approximating such strings by statistical analysis of a huge corpus, facts understood very well in the early stages of inquiry into these topics in the 1950s.

Emergence of unbounded Merge at once provides a kind of ‘language of thought’, an internal system that makes use of conceptual ‘atoms’ (perhaps pre-existent) to construct expressions of arbitrary richness and complexity. It is often argued that another independent language of thought must be postulated, but the arguments for that do not seem to me compelling, a matter that would carry us too far afield.

As a simple matter of logic, there are two kinds of Merge, external and internal. External Merge (EM) takes two objects, say ‘eat’ and ‘apples’, and forms the new object that corresponds to ‘eat apples’. Internal Merge (IM) – often called Move – is the same, except that one of the objects is internal to the other. So applying IM to ‘John saw what’, we form the new internal object corresponding to ‘what John saw what’. At the semantic interface, both occurrences of ‘what’ are interpreted, the first occurrence as an operator and the second as the variable over which it ranges, along the lines I mentioned earlier. At the sensorimotor side, only the first occurrence is pronounced. That illustrates the ubiquitous displacement property of language: items are commonly pronounced in one position but interpreted somewhere else. The displacement property has been regarded in the past, by me in particular, as a curious ‘imperfection’ of human language, a departure from good design. That was an error, however. The property ‘comes free’: some stipulation would be required to block it. One copy must be pronounced, or there would be no indication that the lexical items in it had ever been introduced into the derivation. Failure to pronounce all other
copies follows from third-factor considerations of efficient computation, since it reduces the burden of repeated application of the rules that map internal structures to phonetic form – a very considerable burden when we consider more complex cases. The questions do not arise on the semantic side. As always, there is more to say, but this seems the heart of the matter.

This simple example suggests that the relation of the I-language to the interfaces is asymmetrical. Optimal design yields the right properties at the semantic side, but causes processing problems at the sound side. Those familiar with perceptual models and parsing systems will be aware that some of the major problems they face would be overcome if all copies were pronounced, not just the structurally most prominent one (typically the first one), so that gaps have to be located somehow. In this case, conditions of efficient computation conflict with facilitation of communication, and universally, languages prefer efficient computation. That appears to be true quite generally. If so, it appears that language evolved, and is designed, primarily as an instrument of thought, with externalization a secondary process.

There are other reasons to believe that something like that is true. One is that externalization appears to be at least in part modality-independent, as has been learned from studies of sign language in recent years. The structural properties of sign and spoken language appear to be remarkably similar. Acquisition follows the same course, and neural localization seems to be similar as well. Children raised in bimodal environments – one parent speaking and the other signing – seem to exhibit no preference, treating the two languages just as if they were exposed to Spanish and English. That tends to reinforce the conclusion that language is optimized for the system of thought, with mode of externalization secondary.

There are more general reasons that suggest the same conclusion. The core principle of language, unbounded Merge, must have arisen from some rewiring of the brain, presumably not too long before the ‘great leap forward’, and hence very recently in evolutionary time. Such changes take place in an individual, not a group. The individual so endowed would have had many advantages: capacities for complex thought, planning, interpretation, and so on. The capacity would be transmitted to offspring, coming to dominate a small breeding group. At that stage, there would be an advantage to externalization, so the capacity would be linked as a secondary process to the sensorimotor system for externalization and interaction, including communication – a special case, at least if we invest the term ‘communication’ with some meaning. It is not easy to imagine an account of human evolution that does not assume at least this much, in one or another form. Any additional assumption requires both evidence and rationale, not easy to come by.

Similar proposals have been made by eminent evolutionary biologists. At the international conference on biolinguistics thirty years ago, Salvador Luria was the most forceful advocate of the view that communicative needs
would not have provided ‘any great selective pressure to produce a system such as language’, with its crucial relation to ‘development of abstract or productive thinking’ (Luria, 1974). The same idea was taken up by his fellow Nobel laureate François Jacob, who suggested that ‘the role of language as a communication system between individuals would have come about only secondarily…. The quality of language that makes it unique does not seem to be so much its role in communicating directives for action’ or other common features of animal communication, but rather ‘its role in symbolizing, in evoking cognitive images’, in ‘molding’ our notion of reality and yielding our capacity for thought and planning, through its unique property of allowing ‘infinite combinations of symbols’ and therefore ‘mental creation of possible worlds’ (Jacob, 1982: p. 59). These ideas can be traced back to the cognitive revolution of the seventeenth century, which in many ways foreshadows developments from the 1950s, as I’ve mentioned.

We can, however, go beyond speculation. Investigation of language design can yield evidence on the relation of language to the interfaces. There is, I think, mounting evidence that the relation is asymmetrical in the manner indicated. If so, principled explanation will reduce to efficient computation and satisfaction of semantic interface conditions, with satisfaction of SM conditions an ancillary process.

The picture fits reasonably well with what is understood about the evolution of modern humans. It appears that human brain size reached its current level recently in evolutionary time, perhaps about 100,000 years ago, which suggests to some specialists that ‘human language probably evolved, at least in part, as an automatic but adaptive consequence of increased absolute brain size’ (Striedter, 2006: p. 10). One of the leading specialists, Ian Tattersall, suggests a much earlier date for anatomically modern humans, though behaviourally modern humans are attested much later, perhaps about 50,000–60,000 years ago. The dates are subject to debate, but not in ways that matter much for our purposes. With regard to language, ‘what must have happened’, Tattersall writes, is that after a long – and poorly understood – period of erratic brain expansion and reorganization in the human lineage, something occurred that set the stage for language acquisition. This innovation would have depended on the phenomenon of emergence, whereby a chance combination of pre-existing elements results in something totally unexpected, presumably ‘a neural change … in some population of the human lineage … rather minor in genetic terms, [which] probably had nothing whatever to do with adaptation’ (Tattersall, 2002: pp. 161–2), though like some other exaptations, it conferred advantages, then proliferated. Perhaps it was an automatic consequence of absolute brain size, as Striedter suggests, or perhaps some minor chance mutation. Sometime later – not very long in evolutionary time – came further innovations, perhaps cultural, that led to behaviourally modern humans, the ‘great leap forward’, and the trek from Africa.
We might speculate, then, that the ‘emergence’ was the capacity for recursive generation in its simplest possible form: the operation Merge, linked to atomic conceptual resources to create a ‘language of thought’, perhaps near optimally. That exaptation would indeed have been adaptive, and could well have come to dominate the particular population of the human lineage in which it occurred. The later event would have been the externalization, laying the basis for rich social interchange and the great leap forward. The externalization is a complex affair for the reasons mentioned, and could have taken place in many different ways, either before or after the dispersal of the original population. The result would be what we seem increasingly to find: a computational system efficiently generating expressions that provide the language of thought, and complex and highly varied modes of externalization. For the core system that emerged in the first stage, variation might be limited to ‘macroparameters’ of the kind discussed, for example, in Mark Baker’s *Atoms of Language* (2001): the polysynthesis, head-complement, null-subject, and some other parameters. The varied and complex modes of externalization yield a rich variety of ‘microparameters’, for example, whatever distinguishes the strange way in which ‘can’ (tin receptacle) and ‘can’ (be able) are pronounced in the particular corner of Philadelphia where I grew up. That general picture corresponds reasonably well with what is so far more or less understood about language, including the available partial answers to the ‘what’, ‘how’, and ‘why’ questions.

If this line of argument can be sustained, the problems facing the study of evolution of language can be sharpened. It is necessary to account for (1) the atoms for computation (lexical items, concepts), with their apparently unique human characteristics, and their internal properties and relations; (2) the rewiring of the brain that made unbounded Merge available for generating structured expressions from these atoms; (3) the operations of externalization that map expressions to the sensorimotor interface; (4) whatever else remains in UG after ‘why’ questions have been answered by resort to third-factor principles; (5) interface operations. Problem (1) has to do more broadly with the evolution of human thought, and has barely been addressed, in part because of failure to recognize the specific human properties of these elements. Not enough is known about the brain to formulate (2) in a useful way. The variety and complexity of language appear to reside overwhelmingly in category (3), which would not be surprising if indeed externalization is an ancillary process, relating two systems that are quite distinct: the generation of the semantic interface and sensorimotor systems. An interesting question for the future is to determine to what extent languages ‘solve’ this problem in an optimal way, given the many extraneous conditions that have to be satisfied (for example, the effects of the Norman conquest). Investigation of (4) remains open insofar as inquiry beyond explanatory adequacy is not approaching what appear to be its
limits. At the sensorimotor side, questions (5) are the topic of articulatory
and acoustic phonetics, and their analogues for other modalities, at least
sign. At the semantic interface, the questions remain obscure, for one
reason, because of the difficulty of investigating the systems of thought
without recourse to language, thus at least partially not begging the crucial
questions.

At this point we have to move on to more technical discussion than is
possible here, but I think it is fair to say that there has been considerable
progress in moving towards principled explanation in terms of third-factor
considerations, considerably sharpening our understanding of the specific
properties of language – the core problem of theoretical study of language
since its origins millennia ago, now taking quite new forms.

The quest for principled explanation faces daunting tasks. We can
formulate the goals with reasonable clarity. With each step towards the
goal we gain a clearer grasp of the universals of language. It should be
kept in mind, however, that any such progress still leaves unresolved
problems that have been raised for hundreds of years. Among these are
the question how properties ‘termed mental’ relate to ‘the organical
structure of the brain’, in the eighteenth-century formulation, and the
mysterious problems of the creative and coherent ordinary use of
language, a central problem of Cartesian science, still scarcely even at the
horizons of inquiry.

MIT, Cambridge, Mass., USA

References

Huarte, Juan (1575) Examen de ingenios para las ciencias, Baeza.
Book I, Part IV, Sect. VI, ‘Of Personal Identity’.
Joos, Martin (ed.) (1966) Readings in Linguistics, 4th edn, Chicago: University of
Chicago Press.
Kripke, Saul (1979) ‘A Puzzle about Belief’, in A. Margalit (ed.) Meaning and Use,
Locke, John (1689/1975) An Essay Concerning Human Understanding, Oxford:
Oxford University Press.