# Introduction

During the past half-century, there has been intensive and often highly productive inquiry into human cognitive faculties, their nature and the ways they enter into action and interpretation. Commonly it adopts the thesis that "things mental, indeed minds, are emergent properties of brains," while recognizing that "those emergences are . . . produced by principles that control the interactions between lower level events – principles we do not yet understand" (Mountcastle 1998: 1). The word "yet" expresses the optimism that has, rightly or wrongly, been a persistent theme throughout the period.

The thesis revives eighteenth-century proposals that were put forth for quite compelling reasons: in particular, the conclusion that Newton appeared to have established, to his considerable dismay, that "a purely materialistic or mechanistic physics" is "impossible" (Koyré 1957: 210); and the implications of "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 1975: 541, Book IV, Chapter 3, Section 6). The precedents of the early modern period, and the thinking that lay behind them, merit closer attention than they have generally, in my opinion, received. It is also worth remembering that lack of understanding of "mind/brain interaction" is not the only respect in which progress has been limited since the origin of the modern scientific revolutions. While inquiry into higher mental faculties has achieved a great deal in some areas, the results do not reach the issues that were – sensibly in my view – taken to be at the heart of the problem. Some of these topics are touched on in the following chapters.

One domain in which there has been substantial progress is the study of language, particularly in the past 20 years. Here too, traditional questions remain at the horizon, if even there. My understanding of this work is that it (often implicitly) takes for granted some version of the thesis on mind/brain just quoted, and can reasonably be interpreted as part of psychology or, more broadly, human biology. Some have plausibly

termed it "biolinguistics" (Jenkins 1999). Its topic is particular states of people, mostly their brains: call them "linguistic states." It seeks to unearth the nature and properties of such states, their development and variety, and their basis in innate biological endowment. That endowment appears to determine a "faculty of language" that is a distinctive component of higher mental faculties (as a system, that is, its elements may have all sorts of functions), a "species-property" that is shared among humans to close approximation, over a broad range. The faculty of language is a very recent evolutionary development and, as far as is known, is biologically isolated in crucial respects. Biolinguistic inquiry seeks unification with other approaches to the properties of the brain, in the hope that some day the slash "/" in the phrase "mind/brain" will gain more substantive content. It is concerned not only with the nature and development of linguistic states, but also with the ways they enter into the use of language. Included in principle, sometimes in fact, are the relations of these states to an external medium (production and perception), and their role in thinking and talking about the world and other human actions and interactions. In some domains, particularly with regard to problems of reference and meaning in natural language, the approach seems to me to suggest that considerable rethinking may be in order, for reasons discussed in the following chapters.

It has to be shown, of course, that this "naturalistic" approach is a proper way to investigate phenomena of language, and the use of language. A more ambitious thesis is that it is presupposed (at least tacitly, and sometimes in the face of explicit denial) by constructive work generally in these areas; and that something similar holds in the study of other cognitive faculties. It must also be shown that critiques are misguided, including those that are widespread and influential. I think all of this is rather plausible. The essays that follow, mostly based on talks over the past few years, attempt to provide some reasons for these conclusions, and to sketch some directions that seem to me appropriate and worth exploring.

# 1 New horizons in the study of language

The study of language is one of the oldest branches of systematic inquiry, tracing back to classical India and Greece, with a rich and fruitful history of achievement. From a different point of view, it is quite young. The major research enterprises of today took shape only about 40 years ago, when some of the leading ideas of the tradition were revived and reconstructed, opening the way to what has proven to be very productive inquiry.

That language should have exercised such fascination over the years is not surprising. The human faculty of language seems to be a true "species property," varying little among humans and without significant analogue elsewhere. Probably the closest analogues are found in insects, at an evolutionary distance of a billion years. There is no serious reason today to challenge the Cartesian view that the ability to use linguistic signs to express freely-formed thoughts marks "the true distinction between man and animal" or machine, whether by "machine" we mean the automata that captured the imagination of the seventeenth and eighteenth century, or those that are providing a stimulus to thought and imagination today.

Furthermore, the faculty of language enters crucially into every aspect of human life, thought, and interaction. It is largely responsible for the fact that alone in the biological world, humans have a history, cultural evolution and diversity of any complexity and richness, even biological success in the technical sense that their numbers are huge. A Martian scientist observing the strange doings on Earth could hardly fail to be struck by the emergence and significance of this apparently unique form of intellectual organization. It is even more natural that the topic, with its many mysteries, should have stimulated the curiosity of those who seek to understand their own nature and their place within the wider world.

Human language is based on an elementary property that also seems to be biologically isolated: the property of discrete infinity, which is exhibited in its purest form by the natural numbers 1, 2, 3, . . . Children do not

learn this property; unless the mind already possesses the basic principles, no amount of evidence could provide them. Similarly, no child has to learn that there are three and four word sentences, but no three-and-a half word sentences, and that they go on forever; it is always possible to construct a more complex one, with a definite form and meaning. Such knowledge must come to us from "the original hand of nature," in David Hume's (1748/1975: 108, Section 85) phrase, as part of our biological endowment.

This property intrigued Galileo, who regarded the discovery of a means to communicate our "most secret thoughts to any other person with 24 little characters" (Galileo 1632/1661, end of first day) as the greatest of all human inventions. The invention succeeds because it reflects the discrete infinity of the language that these characters are used to represent. Shortly after, the authors of the Port Royal Grammar were struck by the "marvellous invention" of a means to construct from a few dozen sounds an infinity of expressions that enable us to reveal to others what we think and imagine and feel – from a contemporary standpoint, not an "invention" but no less "marvellous" as a product of biological evolution, about which virtually nothing is known, in this case.

The faculty of language can reasonably be regarded as a "language organ" in the sense in which scientists speak of the visual system, or immune system, or circulatory system, as organs of the body. Understood in this way, an organ is not something that can be removed from the body, leaving the rest intact. It is a subsystem of a more complex structure. We hope to understand the full complexity by investigating parts that have distinctive characteristics, and their interactions. Study of the faculty of language proceeds in the same way.

We assume further that the language organ is like others in that its basic character is an expression of the genes. How that happens remains a distant prospect for inquiry, but we can investigate the genetically-determined "initial state" of the language faculty in other ways. Evidently, each language is the result of the interplay of two factors: the initial state and the course of experience. We can think of the initial state as a "language acquisition device" that takes experience as "input" and gives the language as an "output" – an "output" that is internally represented in the mind/brain. The input and the output are both open to examination: we can study the course of experience and the properties of the languages that are acquired. What is learned in this way can tell us quite a lot about the initial state that mediates between them.

Furthermore, there is strong reason to believe that the initial state is common to the species: if my children had grown up in Tokyo, they

would speak Japanese, like other children there. That means that evidence about Japanese bears directly on the assumptions concerning the initial state for English. In such ways, it is possible to establish strong empirical conditions that the theory of the initial state must satisfy, and also to pose several problems for the biology of language: How do the genes determine the initial state, and what are the brain mechanisms involved in the initial state and the later states it assumes? These are extremely hard problems, even for much simpler systems where direct experiment is possible, but some may be at the horizons of inquiry.

The approach I have been outlining is concerned with the faculty of language: its initial state, and the states it assumes. Suppose that Peter's language organ is in state L. We can think of L as Peter's "internalized language." When I speak of a language here, that is what I mean. So understood, a language is something like "the way we speak and understand," one traditional conception of language.

Adapting a traditional term to a new framework, we call the theory of Peter's language the "grammar" of his language. Peter's language determines an infinite array of expressions, each with its sound and meaning. In technical terms, Peter's language "generates" the expressions of his language. The theory of his language is therefore called a generative grammar. Each expression is a complex of properties, which provide "instructions" for Peter's performance systems: his articulatory apparatus, his modes of organizing his thoughts, and so on. With his language and the associated performance systems in place, Peter has a vast amount of knowledge about the sound and meaning of expressions, and a corresponding capacity to interpret what he hears, express his thoughts, and use his language in a variety of other ways.

Generative grammar arose in the context of what is often called "the cognitive revolution" of the 1950s, and was an important factor in its development. Whether or not the term "revolution" is appropriate, there was an important change of perspective: from the study of behavior and its products (such as texts), to the inner mechanisms that enter into thought and action. The cognitive perspective regards behavior and its products not as the object of inquiry, but as data that may provide evidence about the inner mechanisms of mind and the ways these mechanisms operate in executing actions and interpreting experience. The properties and patterns that were the focus of attention in structural linguistics find their place, but as phenomena to be explained along with innumerable others, in terms of the inner mechanisms that generate expressions. The approach is "mentalistic," but in what should be an uncontroversial sense. It is concerned with "mental aspects of the world," which stand alongside its mechanical, chemical, optical, and

other aspects. It undertakes to study a real object in the natural world – the brain, its states, and its functions – and thus to move the study of the mind towards eventual integration with the biological sciences.

The "cognitive revolution" renewed and reshaped many of the insights, achievements, and quandaries of what we might call "the first cognitive revolution" of the seventeenth and eighteenth century, which was part of the scientific revolution that so radically modified our understanding of the world. It was recognized at the time that language involves "the infinite use of finite means," in Wilhelm von Humboldt's phrase; but the insight could be developed only in limited ways, because the basic ideas remained vague and obscure. By the middle of the twentieth century, advances in the formal sciences had provided appropriate concepts in a very sharp and clear form, making it possible to give a precise account of the computational principles that generate the expressions of a language, and thus to capture, at least partially, the idea of "infinite use of finite means." Other advances also opened the way to investigation of traditional questions with greater hope of success. The study of language change had registered major achievements. Anthropological linguistics provided a far richer understanding of the nature and variety of languages, also undermining many stereotypes. And certain topics, notably the study of sound systems, had been much advanced by the structural linguistics of the twentieth century.

The earliest attempts to carry out the program of generative grammar quickly revealed that even in the best studied languages, elementary properties had passed unrecognized, that the most comprehensive traditional grammars and dictionaries only skim the surface. The basic properties of languages are presupposed throughout, unrecognized and unexpressed. That is quite appropriate if the goal is to help people to learn a second language, to find the conventional meaning and pronunciation of words, or to have some general idea of how languages differ. But if our goal is to understand the language faculty and the states it can assume, we cannot tacitly presuppose "the intelligence of the reader." Rather, this is the object of inquiry.

The study of language acquisition leads to the same conclusion. A careful look at the interpretation of expressions reveals very quickly that from the earliest stages, the child knows vastly more than experience has provided. That is true even of simple words. At peak periods of language growth, a child is acquiring words at a rate of about one an hour, with extremely limited exposure under highly ambiguous conditions. The words are understood in delicate and intricate ways that are far beyond the reach of any dictionary, and are only beginning to be investigated. When we move beyond single words, the conclusion

becomes even more dramatic. Language acquisition seems much like the growth of organs generally; it is something that happens to a child, not that the child does. And while the environment plainly matters, the general course of development and the basic features of what emerges are predetermined by the initial state. But the initial state is a common human possession. It must be, then, that in their essential properties and even down to fine detail, languages are cast to the same mold. The Martian scientist might reasonably conclude that there is a single human language, with differences only at the margins.

As languages were more carefully investigated from the point of view of generative grammar, it became clear that their diversity had been underestimated as radically as their complexity and the extent to which they are determined by the initial state of the faculty of language. At the same time, we know that the diversity and complexity can be no more than superficial appearance.

These were surprising conclusions, paradoxical but undeniable. They pose in a stark form what has become the central problem of the modern study of language: How can we show that all languages are variations on a single theme, while at the same time recording faithfully their intricate properties of sound and meaning, superficially diverse? A genuine theory of human language has to satisfy two conditions: "descriptive adequacy" and "explanatory adequacy." The grammar of a particular language satisfies the condition of descriptive adequacy insofar as it gives a full and accurate account of the properties of the language, of what the speaker of the language knows. To satisfy the condition of explanatory adequacy, a theory of language must show how each particular language can be derived from a uniform initial state under the "boundary conditions" set by experience. In this way, it provides an explanation of the properties of languages at a deeper level.

There is a serious tension between these two research tasks. The search for descriptive adequacy seems to lead to ever greater complexity and variety of rule systems, while the search for explanatory adequacy requires that language structure must be invariant, except at the margins. It is this tension that has largely set the guidelines for research. The natural way to resolve the tension is to challenge the traditional assumption, carried over to early generative grammar, that a language is a complex system of rules, each specific to particular languages and particular grammatical constructions: rules for forming relative clauses in Hindi, verb phrases in Swahili, passives in Japanese, and so on. Considerations of explanatory adequacy indicate that this cannot be correct.

The central problem was to find general properties of rule systems that can be attributed to the faculty of language itself, in the hope that the residue will prove to be more simple and uniform. About 15 years ago, these efforts crystallized in an approach to language that was a much more radical departure from the tradition than earlier generative grammar had been. This "Principles and Parameters" approach, as it has been called, rejected the concept of rule and grammatical construction entirely: there are no rules for forming relative clauses in Hindi, verb phrases in Swahili, passives in Japanese, and so on. The familiar grammatical constructions are taken to be taxonomic artifacts, useful for informal description perhaps but with no theoretical standing. They have something like the status of "terrestrial mammal" or "household pet." And the rules are decomposed into general principles of the faculty of language, which interact to yield the properties of expressions.

We can think of the initial state of the faculty of language as a fixed network connected to a switch box; the network is constituted of the principles of language, while the switches are the options to be determined by experience. When the switches are set one way, we have Swahili; when they are set another way, we have Japanese. Each possible human language is identified as a particular setting of the switches – a setting of parameters, in technical terminology. If the research program succeeds, we should be able literally to deduce Swahili from one choice of settings, Japanese from another, and so on through the languages that humans can acquire. The empirical conditions of language acquisition require that the switches can be set on the basis of the very limited information that is available to the child. Notice that small changes in switch settings can lead to great apparent variety in output, as the effects proliferate through the system. These are the general properties of language that any genuine theory must capture somehow.

This is, of course, a program, and it is far from a finished product. The conclusions tentatively reached are unlikely to stand in their present form; and, needless to say, one can have no certainty that the whole approach is on the right track. As a research program, however, it has been highly successful, leading to a real explosion of empirical inquiry into languages of a very broad typological range, to new questions that could never even have been formulated before, and to many intriguing answers. Questions of acquisition, processing, pathology, and others also took new forms, which have proven very productive as well. Furthermore, whatever its fate, the program suggests how the theory of language might satisfy the conflicting conditions of descriptive and explanatory adequacy. It gives at least an outline of a genuine theory of language, really for the first time.

Within this research program, the main task is to discover and clarify the principles and parameters and the manner of their interaction, and to extend the framework to include other aspects of language and its use. While a great deal remains obscure, there has been enough progress to at least consider, perhaps to pursue, some new and more far-reaching questions about the design of language. In particular, we can ask how good the design is. How close does language come to what some super-engineer would construct, given the conditions that the language faculty must satisfy?

The questions have to be sharpened, and there are ways to proceed. The faculty of language is embedded within the broader architecture of the mind/brain. It interacts with other systems, which impose conditions that language must satisfy if it is to be usable at all. We might think of these as "legibility conditions," in the sense that other systems must be able to "read" the expressions of the language and use them as "instructions" for thought and action. The sensorimotor systems, for example, have to be able to read the instructions having to do with sound, that is the "phonetic representations" generated by the language. The articulatory and perceptual apparatus have specific design that enables them to interpret certain phonetic properties, not others. These systems thus impose legibility conditions on the generative processes of the faculty of language, which must provide expressions with the proper phonetic form. The same is true of conceptual and other systems that make use of the resources of the faculty of language: they have their intrinsic properties, which require that the expressions generated by the language have certain kinds of "semantic representations," not others. We may therefore ask to what extent language is a "good solution" to the legibility conditions imposed by the external systems with which it interacts. Until quite recently this question could not seriously be posed, even formulated sensibly. Now it seems that it can, and there are even indications that the language faculty may be close to "perfect" in this sense; if true, this is a surprising conclusion.

What has come to be called "the Minimalist Program" is an effort to explore these questions. It is too soon to offer a firm judgment about the project. My own judgment is that the questions can now profitably be placed on the agenda, and that early results are promising. I would like to say a few words about the ideas and the prospects, and then to return to some problems that remain at the horizons.

The minimalist program requires that we subject conventional assumptions to careful scrutiny. The most venerable of these is that language has sound and meaning. In current terms, that translates in a natural way to the thesis that the faculty of language engages other

systems of the mind/brain at two "interface levels," one related to sound, and the other to meaning. A particular expression generated by the language contains a phonetic representation that is legible to the sensorimotor systems, and a semantic representation that is legible to conceptual and other systems of thought and action.

One question is whether there are levels other than the interface levels: Are there levels "internal" to the language, in particular, the levels of deep and surface structure that have been postulated in modern work? (see, for example, Chomsky 1965; 1981a; 1986). The minimalist program seeks to show that everything that has been accounted for in terms of these levels has been misdescribed, and is as well or better understood in terms of legibility conditions at the interface: for those of you who know the technical literature, that means the projection principle, binding theory, Case theory, the chain condition, and so on.

We also try to show that the only computational operations are those that are unavoidable on the weakest assumptions about interface properties. One such assumption is that there are word-like units: the external systems have to be able to interpret such items as "Peter" and "tall." Another is that these items are organized into larger expressions, such as "Peter is tall." A third is that the items have properties of sound and meaning: the word "Peter" begins with closure of the lips and is used to refer to persons. The language therefore involves three kinds of elements:

- the properties of sound and meaning, called "features";
- the items that are assembled from these properties, called "lexical items"; and
- the complex expressions constructed from these "atomic" units.

It follows that the computational system that generates expressions has two basic operations: one assembles features into lexical items, the second forms larger syntactic objects out of those already constructed, beginning with lexical items.

We can think of the first operation as essentially a list of lexical items. In traditional terms, this list – called the lexicon – is the list of "exceptions," arbitrary associations of sound and meaning and particular choices among the inflectional properties made available by the faculty of language that determine how we indicate that nouns and verbs are plural or singular, that nouns have nominative or accusative case, and so on. These inflectional features turn out to play a central role in computation.

Optimal design would introduce no new features in the course of computation. There should be no indices or phrasal units and no bar levels (hence no phrase-structure rules or X-bar theory; see Chomsky