Some simple evo-devo theses: how true might they be for language?

Study of evolution of some system is feasible only to the extent that its nature is understood. That seems close to truism. A sensible approach is to begin with properties of the system that are understood with some confidence and seek to determine how they may have evolved, then turning to less obvious properties and asking what additional problems they pose for inquiry into evolution. I’ll try to outline such a course for language, keeping to a sketch of general directions.

I will also mention some analogies between what some biologists call “the Evo Devo revolution” and ideas that have been lurking in the background of “biolinguistics” since its origins about half a century ago, and that have been pursued more intensively in recent years. The analogies have been suggestive in the past, and might turn out to be more than that in the years ahead.

The term “biolinguistics” was proposed in 1974 by Massimo Piattelli as the topic for an international conference he organized, held at MIT, bringing together evolutionary biologists, neuroscientists, linguists, and others concerned with language and biology, one of many such initiatives. A primary focus of the discussions was the extent to which apparent principles of language are unique to this cognitive system, plainly one of “the basic questions to be asked from the biological point of view” and crucial for the study of development of language in the individual and its evolution in the species.

The term “language” as used in this context means internal language, sometimes called “I-language,” 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 conceptual/intensional or semantic systems that use linguistic expressions for thought and for organizing action, and the sensorimotor systems that externalize expressions in production and assign them to sensory data in perception. Languages so construed are particular instantiations of some genetically-determined format, which we can call the language faculty, adapting a traditional term to this framework. Certain configurations are possible human languages, others are not, and a primary concern of any theory of human language is to establish the distinction between the two categories.

At the time of the 1974 conference, it seemed that the language faculty must be rich, highly structured, and substantially unique to this cognitive system. In particular, that conclusion followed from considerations of language acquisition. The only plausible idea seemed to be that language acquisition is rather like theory construction. Somehow, the child reflexively categorizes certain sensory data as linguistic, not a trivial achievement in itself, and then uses the constructed linguistic experience as evidence for a theory that generates an infinite variety of expressions, each of which contains the information about sound, meaning, and structure that is relevant for the myriad varieties of language use.

To give a few of the early illustrations for concreteness, the internal theory that those of us in this room more or less share determines that the sentence “Mary saw the man walking to the bus station” is three-ways ambiguous, but the ambiguities are resolved if we ask “which bus station did Mary see the man walking to?” The explanation appears to rely on computationally plausible principles of minimal search, for which there is a good deal of independent evidence. The phrase “which bus station” raises from the position in which its semantic role is determined and is reinterpreted as an operator taking scope over a variable in its original position, so the sentence means, roughly, “for which x, x a bus station, Mary saw the
man walking to x”; the variable is silent in the phonetic output, but must be there for interpretation. Only one of the underlying interpretations permits the operation, by virtue of the minimal search conditions, so the ambiguity is resolved in the interrogative.

To take a second example, consider the sentence “John ate an apple.” We can omit “an apple,” yielding “John ate,” which we understand to mean “John ate something or other, unspecified. Now consider “John was too angry to talk to Bill.” We can omit “Bill,” yielding “John was too angry to talk to,” which, by analogy to “John ate,” would be expected to mean that John was so angry that he wouldn’t talk to someone or other. But it doesn’t mean that: rather, that John is so angry that someone or other won’t talk to him, John. In this case, the explanation lies in the fact that the phrase “too angry to talk to,” with the object missing, actually has an operator-variable structure based on movement of a phonetically invisible operator meeting the same conditions as in the “bus station” example. In this case, the operator has no content and is silent, yielding an open sentence with a free variable, hence a predicate. Again, there is substantial independent evidence supporting the conclusion, 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 has been understood from the earliest work in generative grammar that determination of the grammatical status of a sentence, or efforts to approximate what appears in a corpus, are of only marginal interest. The language that every person quickly masters relies on inner resources to generate internal expressions that yield information of the kind just illustrated, only very partially revealed in a corpus of data, no matter how huge.

Even the most elementary considerations yield the same conclusions. In the earliest work in generative grammar 50 years ago, it was assumed that phonetic units can be identified in a corpus, and that words can be detected by study of transitional probabilities (which, surprisingly, turns out to be false, recent work has shown). We also proposed methods with an information-theoretic flavor to assign such words to categories. But it was evident that even the simplest lexical items raise fundamental problems for analytic procedures of segmentation, classification, statistical analysis, and the like. A lexical item is identified by phonological elements that determine its sound along with morphological elements that determine its meaning. But neither the phonological nor morphological elements have the “beads-on-a-string” property required for computational analysis of a corpus. Furthermore, rather as in the case of the sentences I gave as examples, even the simplest words in many languages have phonological and morphological elements that are silent. The elements that constitute lexical items find their place in the generative procedures that yield the expressions, but cannot be detected in the physical signal. For that reason, it seemed – and seems – that the language acquired must have the basic properties of an internalized explanatory theory. These are elementary and quite general properties that any account of evolution of language must deal with.

Quite generally, construction of theories must be guided by what Charles Sanders Peirce a century ago called an “abductive principle,” genetically determined, which “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 the evidence. For language development, the format that limits admissible hypotheses must, furthermore, be highly restrictive, given the empirical facts of rapidity of acquisition and convergence among individuals. The conclusions about the specificity and richness of the language faculty
seemed to follow directly. Plainly such conclusions pose serious barriers to inquiry into how the faculty might have evolved, matters discussed repeatedly, and inconclusively, at the 1974 conference.

A few years later, a new approach crystallized that suggested ways in which these barriers might be overcome. This “Principles and Parameters” (P&P) approach was based on the idea that the format consists of invariant principles and a “switch-box” of parameters that can be set to one or another value on the basis of fairly elementary experience. A choice of parameter settings determines a language. The approach largely emerged from intensive study of a range of languages, but it was also suggested by an analogy to early evo-devo discoveries, specifically François Jacob’s account of how slight changes in regulatory mechanisms can 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. The approach has been pursued with considerable success, with many improvements and revisions along the way. One illustration is Mark Baker’s demonstration, in his book *Atoms of Language*, that languages that appear on the surface to be radically different, such as Mohawk and English, turn out to be remarkably similar when we abstract from the effects of a few choices of values for parameters with a hierarchic organization that he argues to be universal, hence the outcome of evolution of language.

The approach stimulated highly productive investigation of languages of great typological variety, and also reinvigorated neighboring fields, particularly the study of language acquisition, reframed as inquiry into setting of parameters in the early years of life, with very fruitful results. The approach also provided a new perspective to undermine the long-standing though implausible belief that languages can “differ from each other without limit and in unpredictable ways,” in the words of a prominent theoretician summarizing received opinion in the mid-1950s, with some exaggeration but not too much. Similar views were familiar in biology as well. Thus until quite recently it appeared that variability of organisms is so great as to constitute “a near infinitude of particulars which have to be sorted out case by case” (molecular biologist Gunther Stent), conceptions now significantly modified by evo-devo discoveries about organizing principles, deep homologies, and conservation of fundamental mechanisms of development, perhaps most famously hox genes.

The P&P approach also removed a major conceptual barrier to the study of evolution of language. With the divorce of principles of language from acquisition, it no longer follows that the format that “limits admissible hypotheses” must be rich and highly structured to satisfy the empirical conditions of language acquisition. That might turn out to be the case, but it is no longer an apparent conceptual necessity.

Here too research programs within linguistics had certain analogies to the evo-devo revolution, including the discovery, quoting Jacob and others, 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.” Evidently, development of language in the individual must involve three factors: First, genetic endowment, which sets limits on the languages attained; second, external data, which selects one or another language within a narrow range; and third, principles not specific to the language faculty. The third factor principles have the flavor of the architectural and developmental constraints that enter into all facets of growth and evolution. Among these are principles of efficient computation, such as those I mentioned. These would be expected to be of particular significance for generative systems such as the internalized
language. Insofar as the third factor can be shown to be operative in the design of the language faculty, the
task of accounting for its evolution is correspondingly eased.

Recent inquiry into these topics has come to be called “the minimalist program,” but it should be noted that
the program is both traditional and pretty much theory neutral. The serious study of language has always
sought to discover what is distinctive to this cognitive faculty, hence implicitly abstracting from third factor
effects. And whatever one’s beliefs about design of language may be, the questions of the minimalist
research program arise.

Let’s turn to the approach I suggested at the outset: beginning with the properties of language that are
understood with some confidence. The internal language, again, is a computational system that generates
infinitely many internal expressions, each an array of instructions to the interface systems, sensorimotor
and semantic. To the extent that third factor conditions function, the language will be efficiently designed
to satisfy conditions imposed at the interface. We can regard an account of some linguistic phenomena as
principled insofar as it derives them by efficient computation satisfying interface conditions.

Any generative system, natural or invented, is based on an operation that takes structures already formed
and combines them into a new structure. Call it Merge. Operating without bounds, Merge yields a discrete
infinity of structured expressions. The only alternatives are, effectively, notational variants. Hence
unbounded Merge must be part of the genetic component of the language faculty, a product of the evolution
of this “cognitive organ.”

Notice that the conclusion holds whether such recursive generation is unique to the language faculty or
found elsewhere. If the latter, there still must be a genetic instruction to use unbounded Merge to form
structured linguistic expressions satisfying the interface conditions. Nonetheless, it is interesting to ask
whether this operation is language-specific. We know that it is not. The classic illustration is the system of
natural numbers. That brings up a problem posed by Alfred Russell Wallace 125 years ago: in his words,
the “gigantic development of the mathematical capacity is wholly unexplained by the theory of natural
selection, and must be due to some altogether distinct cause,” if only because it remained unused. One
possibility is that it is derivative from language. It is not hard to show that if the lexicon is reduced to a
single element, then unbounded Merge will yield arithmetic. Speculations about the origin of the
mathematical capacity as an abstraction from linguistic operations are familiar, as are criticisms, including
apparent dissociation with lesions and diversity of localization. The significance of such phenomena,
however, is far from clear; they relate to use of the capacity, not its possession. For similar reasons,
dissociations do not show that the capacity to read is not parasitic on the language faculty.

Suppose the single item in the lexicon is a complex object, say some visual array. Then Merge will yield a
discrete infinity of visual patterns, but this is simply a special case of arithmetic. The same would be true if
we add a recursive operation – another instance of unbounded Merge – to form an infinite lexicon, on the
model of some actual (if rather trivial) lexical rules of natural language. This is still just a more elaborate
form of arithmetic, raising no new issue. Similar questions might be asked about the planning systems
investigated by George Miller and associates 45 years ago. If these and other cases fall under the same
general rubric, then unbounded Merge is not only a genetically determined property of language, but also
unique to it.
Either way, evolution of language required some innovation to provide instructions for unbounded Merge, forming structured expressions accessible to the two interface systems. There are many proposals involving precursors with Merge bounded: an operation to form two-word expressions from single words to reduce memory load for the lexicon, then another operation to form three-word expressions, etc. There is no evidence for this, and no obvious rationale either, since it is still necessary to assume that at some point unbounded Merge appears. Hence the assumption of earlier stages seems superfluous. The same issue arises in language acquisition. The modern study of the topic began with the assumption that the child passes through a two-word state, etc. Again the assumption lacks a rationale, because at some point unbounded Merge must appear. Hence the capacity must have been there all along even if it only comes to function at some later stage. There does appear to be evidence for that conclusion: namely, observation of what children produce. But that carries little weight. It has been shown long ago that what children understand at the early stages far exceeds what they produce, and is quite different in character as well. At the telegraphic speech stage of production, for example, children understand normal speech with the function words in the right places but are baffled by telegraphic speech, as shown by experimental work of Lila Gleitman and associates 40 years ago. Hence for both evolution and development, there seems to be little reason to suppose that there were precursors to unbounded Merge.

Suppose X is merged to Y. Maximally efficient computation will leave X and Y unchanged (the No-Tampering Condition). Plainly, either X is external to Y or is part of Y: external and internal Merge, respectively, the latter sometimes called Move. A well-designed language, lacking arbitrary stipulations, will have both cases. Internal Merge yields the familiar phenomenon of displacement, as in the cases I gave earlier: say a question of the form “what did John see,” with two occurrences of “what,” one pronounced in sentence-initial position, the other deleted by phonological rules mapping to the sensorimotor interface. The full internal expression is interpreted at the semantic interface as an operator-variable construction, with “what” given the same semantic role it has when it is not displaced, as in “who saw what.” In a well-designed language, the two kinds of Merge will have different interface properties. That appears to be true. They correlate with the well-known duality of semantics that has been studied from various points of view. External Merge yields argument structure: agent, patient, goal, predicate, etc. Internal Merge yields discourse-related properties such as topic and old information, scope, etc.

Notice that all of these are elementary properties of optimal Merge, and with quite broad empirical support. They are, therefore, properties to be explained by an account of the evolution of language. They follow from the fact that Merge become available at some point in the evolutionary process, and the assumption that third factor properties of efficient computation enter into language growth (“acquisition”).

Another question is whether the relation of language to the interface systems is symmetrical. At the 1974 symposium, a number of participants suggested that it is not; rather that the primary relation is to the semantic interface, to systems of thought. 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.” 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.” These ideas trace back to the cognitive revolution of the 17th century, which in many ways foreshadows developments from the 1950s.

Generation of expressions to satisfy the semantic interface yields a “language of thought.” If the assumption of asymmetry is correct, then the earliest stage of language would have been just that: a language of thought, used internally. It has been argued that an independent language of thought must be postulated. I think there are reasons for skepticism, but that would take us too far afield.

The empirical question of asymmetry can be approached from the study of existing languages. We can seek evidence to determine whether they are optimized to satisfy one or the other interface system. There is, I think, mounting evidence that the thought systems are indeed primary in this respect, as Luria and Jacob speculated. We have just seen one illustration, in fact: the properties of Internal Merge. The No-Tampering Condition entails that the outcome should include the initial and final occurrences, and all intermediate occurrences. This is correct at the semantic interface; I mentioned a simple case, but it is true far more generally, in quite interesting ways, a phenomenon called “reconstruction.” It is, however, not true at the sensorimotor interface, where all but the final position are deleted (with marginal exceptions not relevant here). Why should this be? Here conditions of computational efficiency and of ease of communication are in conflict. Computational efficiency yields the universally attested facts: only the final position of Internal Merge is pronounced. But that leads to comprehension problems. For parsing programs, and perception, major problems are to locate the “gaps” associated with the element that is pronounced, problems that would largely be overcome if all occurrences were pronounced. The issue does not arise at the semantic interface. The conflict between computational efficiency and ease of communication appears to be resolved, universally, in favor of computational efficiency to satisfy the semantic interface, lending support to speculations about its primacy in language design.

Perhaps related are discoveries about sign languages in recent years, which provide substantial evidence that externalization of language is modality-independent. There are striking cases of invention of sign languages by deaf children exposed to no signing and by a community of deaf people brought together very recently, who spontaneously developed a sign language. In the known cases, sign languages are structurally very much like spoken languages, and follow the same developmental patterns from the babbling stage to full competence. They are also distinguished sharply from the gestural systems of the signers, even when the same gesture is used both iconically and symbolically. Laura Pettito and her colleagues have studied children raised in bimodal (signing-speaking) homes, and have found no preferences or basic differences. Her own conclusion is that even “sensitivity to phonetic-syllabic contrasts is a fundamentally linguistic (not acoustic) process and part of the baby’s biological endowment,” and that the same holds at higher levels of structure. Imaging studies lend further support to the hypothesis that “there exists tissue in the human brain dedicated to a function of human language structure independent of speech and sound,” in her words. Studies of brain damage among signers has led to similar conclusions, as has comparative work by Tecumseh Fitch and Marc Hauser indicating, they suggest, that the sensiromotor systems of earlier hominids were recruited for language but perhaps with no special adaptation.

Comparative work on the second interface, systems of thought, is of course much harder. There do, however, seem to be some critical differences between human conceptual systems and symbolic systems of other animals. Even the simplest words and concepts of human language and thought lack the relation to mind-independent entities that has been reported for animal communication: representational systems based 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 Randy Gallistel's introduction to a volume of papers on animal communication. The symbols of human language and thought are quite different, matters explored in interesting ways by 17th-18th century British philosophers, developing ideas that trace back to Aristotle. There appears to be no reference relation in human language and thought, no relation between an internal symbol and a mind-independent object. What we take to be a river, or a person, or a tree, or water turns out not to be identifiable as a physical construct of some kind. Rather, these are creations of what 17th century investigators called the "cognoscitive powers" that provide us with rich means to refer to the outside world from certain perspectives, but are individuated by mental operations that cannot be reduced to a "peculiar nature belonging" to the thing we are talking about, as Hume summarized a century of inquiry. In this regard, internal conceptual symbols are like the phonetic units of mental representations, such as the syllable [ba]; every particular act externalizing this mental entity yields a mind-independent entity, but there is no mind-independent construct that corresponds to the syllable. Words and concepts appear to be similar in this regard, even the simplest of them. These properties seem to be unique to human language and thought.

If I understand the professional literature correctly, it is reasonable to suppose that fairly recently, not too long before about 50,000 years ago, there was an emergence of creative imagination, language and symbolism generally, mathematics, interpretation and recording of natural phenomena, intricate social practices and the like, yielding what Wallace called “man’s intellectual and moral nature,” now sometimes called “the human capacity.” It is commonly assumed that the faculty of language is essential to the human capacity. In a review of current thinking about these matters, Ian Tattersall writes that he is “almost sure that it was the invention of language” that was the “sudden and emergent” event that was the “releasing stimulus” for the appearance of the human capacity in the evolutionary record -- the “great leap forward” as Jared Diamond called it, the result of some genetic event that rewired the brain, allowing for the origin of modern language with the rich syntax that provides modes of expression of thought, a prerequisite for social development and the sharp changes of behavior that are revealed in the archaeological record, and presumably the occasion for the trek from Africa, where otherwise modern humans had apparently been present for hundreds of thousands of years. The dispersion of humans over the world must post-date the evolution of language, since there is no detectable difference in basic language capacity among contemporary humans. Like Luria and Jacob, Tattersall takes language to be “virtually synonymous with symbolic thought,” implying that externalization is a secondary phenomenon, ideas that I think are supported by internal linguistic evidence, as I mentioned.

Putting these thoughts together, we can suggest what seems to be the simplest speculation about the evolution of language. In some small group from which we all descend, a rewiring of the brain took place yielding the operation of unbounded Merge, applying to concepts with properties of the kind I mentioned. Third factor principles enter to yield an infinite array of structured expressions with interpretations of the kind illustrated: duality of semantics, operator-variable constructions, unpronounced elements with substantial consequences for interpretation and thought, etc. The individual so rewired had many advantages: capacities for complex thought, planning, interpretation, and so on. The capacity is transmitted to offspring, coming to predominate. At that stage, there would be an advantage to externalization, so the capacity might come to 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.
Assuming so, what further properties of language require an evolutionary account? That depends on how far one can proceed in giving a principled account of properties of language, in the sense mentioned earlier: showing that they derive from interface conditions, primarily the semantic interface, by third factor properties of efficient computation and the like. If all properties of language could be given principled explanation, then we would conclude that language is perfectly designed to satisfy semantic conditions, and that the mapping to the sensorimotor interface – phonology and morphology and probably more – is a maximally efficient means to convert syntactically generated expressions to a form accessible to the interface. That is too much to expect, but recent work seems to me to show that the ideal is considerably less remote than would have been imagined not long ago. If so, we may be able to gain new insights into evolution and development of language by inquiry into its fundamental design.