The Evolution of Language: a Linguistic and Internalist Perspective

I needn't remind this audience that nowhere else are challenges to evolutionary explanation more pointed than in the case of human language. Today I would like to talk about just two of these challenges, both in the context of my own two-fold evolutionary lineage: while I spent the first decade of my scientific career on the nitty-gritty details of *Drosophila* and evolutionary population genetics, I've spent the last two decades on computational linguistics. So there will be two parts to my talk. In Part I, I would like to illustrate how talk about the evolution of language can be illuminated by drilling down into the details of evolutionary population genetics, the 'auto mechanics' underpinning evolutionary explanation. Here I will focus on a part of the FOXP2 evolutionary story. My conclusion here is that inverse evolutionary inference – drawing conclusions about past selection from current data – is possibly much, much harder than one might have thought. But, wait, there's a bright side. As for computation and language, since my job is to stand in for Marc Hauser, in Part II, I would like to unpack some of the details about the Fitch-Hauser-Chomsky view, about what they dubbed the "narrow faculty of language" in order to see if we can further bridge the gap between us and other species. My conclusion here will be that there's even less of a gap between us and other species than might have otherwise been suspected even five years ago. So here there glad tidings, quite in line with the flurry of recent results demonstrating that many formerly 'human-only' abilities can be found in other species, from everything like rhythmic entrainment in parrots to speaker-independent phonetic perception in zebra finches.

This is especially good news because one enormous source of skepticism regarding modern linguistic theory is that the structures that universal grammar supposes could simply not have arisen by natural selection given the short time period involved – there's just too much detailed 'stuff' that must evolve and be encoded in the genome as UG. Not only is this true of the linguistic theories of 50 years ago, as pictured here, in 1957's Syntactic Structures [SLIDE of syntactic structures] with its detailed optional or obligatory rules and transformations – take a look at the baroque look of transformation T12 for instance – the same holds in spades for its 1980s replacement, so-called "Government and Binding" theory. This tried to break down Universal Grammar into a set of several dozen modular 'constraint systems', like this one [Picture of Pappi]. While this latter approach as shown here is demonstrably more successful in accounting for the learnability and diversity of language than previous accounts, it too asserts that language is cognitively *sui generis*. This detailed Universal Grammar poses a severe evolutionary puzzle. As it appears that the faculty of language emerged recently and rapidly as measured against evolutionary time – 100 generations since the fall of the Roman Republic, and roughly 20000 generations back to the estimated split with Neandertals – there seems to have been too little time for evolution by natural selection to have built all these highly detailed components on the right-hand side of this picture, some that non-linguists surely don't even recognize – like the 'ECP' ('exceptional case marking', why we say "John believed her to be innocent" rather than "John believed she to be innocent"), or the 'case filter' (eg. Why have to say, I am proud *of* John, not I am proud John). In fact, the more modular and highly detailed the proposed components, the worse the evolutionary burden becomes.

How can we escape this evolutionary dilemma, what has been dubbed "Darwin's problem"? (By Cedric Boecx, Norbert Hornstein, and others.) The relatively short time scale available for the evolution of language suggests that the linguistic specificity of the language faculty as envisaged by the 'old style' generative accounts must be a mirage. We're left we the familiar 'evolutionary bricolage' picture: The faculty of language must be an adventitious combination of operations and principles scavenged from cognition and computation in general with possibly small adventitious additions. On this view, the faculty of language is 'almost' the application of generative cognitive mechanisms to the problem of language, with one or two innovations on top of a mostly pre-existing substrate that a 50,000-100,000 year evolutionary time frame permits. And here I will suggest that by dissecting apart the core operation of the basic combinatorial machinery of syntax, the machinery that builds larger pieces of structure out of smaller bits, we can arrive at a solution where the language faculty is what one gets by marrying these already existing cognitive abilities with a single, small twist – for want of a better term, I will call it simply "Original Sin" – to general principles of cognition and computation that we seem share with other animals, as exemplified by vocal learning songbirds, among other species.

To begin, Part I: why is it so hard in general to study evolution by natural selection? Here I am reminded of my *Drosophila* days. Apart from *E. coli* or, say, yeast, I think it's fair to say that *Drosophila* must be one of the best-studied organisms on the planet. And so we know a lot about *Drosophila*. Or do we? After a year in the lab running electrophoretic gels to ferret out genetic variation and its evolutionary consequences, each summer we would repair ourselves to the sunnier locales of Death Valley, California, since my advisor prided his collecting skills as a naturalist. So we'd hike out to snare Drosophila. It's easy: you use rotten bananas, toss them on a heap on the sand, and, bingo, after a while, you've got a swarm of fruitflies. Then we'd do tricky catch-and-release experiments to see how far the flies might be able migrate in a single day – sometimes, they'd show up 10 miles away. And therein lies the rub. While we know almost everything you would want know about *Drosophila* in the laboratory, we know practically nothing about its *actual* life history in the wild. Where do *Drosophila* live? In trees or bushes? How do *Drosophila* mate? How many larval offspring does a mated female produce, on average? What are the migration patterns of the flies? And so on and so on – in short, *every* detail one would want to know to calculate the *actual* reproductive fitness – age-specific reproductive and mortality data – isn't known. In short, we don't know much beyond banana peels about the *actual* evolutionary story regarding *Drosophila.* And this is by no means the exception. It's the rule.

The study of evolutionary processes and, in particular, the genetics of the evolutionary process must confront special difficulties in both the conceptual and the methodological aspects of research. Indeed, these problems are so severe that evolutionary inference very nearly borders on what mathematicians would say is an "ill posed problem." Why? First, conceptually, unlike for molecular, cellular, and developmental biology, there is no basic mechanism that evolutionists are attempting to elucidate. There is no single cause of the evolutionary change in the properties of members of a species. Natural selection may be involved but so are random events, patterns of migration and interbreeding, mutational events, and horizontal transfer of genes across species boundaries. The change in each character of each species is a consequence of a particular mixture of these causal pathways. Besides selection, other forces like migration and demographic changes, chance events, and horizontal gene transfer from species to species may have a larger role to play – again depending on the case at hand. A concrete example comes from what may perhaps be the most thorough demonstration of selection in natural populations, the 30-year study of two of Darwin's finches (*Geospiza fortis*, the medium ground finch and *Geospiza scandens*, the cactus finch) by P. and B.R. Grant. (2002). While some changes in bill shape and body size were predictable from measured reproductive fitness differences, changes over the entire 30-year study were unpredictable from fitness differences and were the consequence of repeated unforeseen hybridizations between the species, i.e., “selective gains as a result of nonrandom recruitment to the adult population,” with the Grants concluding that “evolution of a population is contingent upon environmental change, which may be highly irregular, as well as on its demography and genetic architecture” such that if one had focused on selection, “the phenotypic states of both species at the end of the 30-year study could not have been predicted at the beginning” (2002:709). (Grant PR, Grant BR., 2002. Unpredictable evolution in a 30-year study of Darwin’s finches. *Science* 296:707–11.)

Further, processes of change and divergence of species are usually extremely slow compared not only to the lifetime of an investigator but of science as an institution. This low speed is, in turn, a consequence of the weakness of most evolutionary forces most of the time. The ‘weakness’ of selection seems to be to be often under-appreciated in discussion about the evolution of language and how to model it. It is now generally agreed that selection differences in nature for most of the gene variants segregating at most loci are likely to be of the order of 10^{-3} or less – that is, a less than 1 part in a thousand difference in fitness. That is: roughly a difference of 1 out of 1000, or 0.1%, where one individual produces 1000 offspring, another 999. We’ll see in a minute that these values are sometimes simply passed over relevant published results. This means that there are many other ‘forces’ – migration, demographic structure, and the like, that can readily mimic the effects of natural selection, i.e., allele frequency changes, so it can be difficult to sort out just *which* of these is responsible. Selection can be easily overwhelmed, and there’s an inherent signal-to-noise problem in ‘detecting’ the hand of selection.

We will see that this has implications for the very real case of FOXP2, but to consider one simplified example of how the powerful these alternative forces can be, there’s the famous red state-blue state problem. (You might not have the same problem in the EU, but there’s always Greece, of course.) Consider the following imaginary situation: suppose we have two “states” in the US. The first state is located in the middle of the US, let’s call it “Kansas”. The second state is on the East coast of the US, let’s call it “New York.” Now, we suppose that the people in both states share a lot of the same properties: they both drive Chevrolets, drink Budweiser beer, vote Republican and so forth. But the two states are separated by a vast gulf of territory. In time, the people in New York might slowly drift away and start to do all sorts of things differently from their Kansas counterparts: drink white wine, eat brie, drive Saabs, vote for Democrats. Worried, the people in Kansas would like to every so often send enough Kansas-type people to New York to prevent this disastrous slide into moral depravity. The question is: how many people would they have to send to New York every generation, on average, to prevent any such divergence? The answer might surprise you. It is astonishingly low: as few as 1 or 2 people per generation. That tells you how easily migration can ‘swamp’ other observed changes in gene frequencies. (And this is precisely what we don’t know about Drosophila.)

So the methodological apparatus for reconstructing past events from present organisms can be very tricky. It is meant to estimate the actual magnitude of forces of natural selection and random events that have led to genetic differences between closely related species or to genetic differences between populations of the same species. It involves a mixture of the mathematical theory of population genetics and statistical theory. It begins with a complex mathematical apparatus that is designed to carry the state of a population forward in time from some initial condition. It predicts rates of genetic change from an initial state and possible equilibria that will result from selection, mutation, migration, and recombination. This must be a stochastic, rather than a deterministic, theory to account for random changes that result from genetic drift in finite populations, so that the form of the prediction is not a unique state at future time, but a probability distribution of states. That's the 'standard' "Modern synthesis" wedding Mendelian genetics to evolutionary change, as developed by R.A. Fisher, Sewall Wright, and others during the 1930s. Going forward from the past to the present, The Modern Synthesis 'auto mechanics' works.

However, that is not the situation we find ourselves in! A probabilistic theory is needed that can *reverse* the deductions of the first theory and infer *backwards* in time from a particular observed state at the present – the DNA sequence data we collect about FOXP2 from different species – to the most likely dynamical forces that have led to the actual present situation. But a difficulty arises here, the issue of 'ill-posed problem' that was referred to earlier. A dynamical theory that predicts the present state generally requires that we know not only the nature and magnitude of the forces that have operated, but also the initial condition and how long the process has been in operation. That means that if we wish to use a backward inference from the present state to estimate the forces that have operated, we would need to know the initial condition and how long the process has been going on as well as assuming that the forces have not changed during the process.

But this is precisely what we cannot know! *Either* we assume that we know the forces, in which case we can make probability statements about the initial conditions, *or else* we assume that we know the initial conditions, in which case we can make estimates of the forces that have led to the present. *We cannot do both*.

It's like solving for three variables, but having only 1 equation instead of three. There is one way out of this dilemma, a 'trick' that is not usually explicitly mentioned by most texts or papers in evolutionary population biology. If the evolutionary process has gone on for a sufficiently long time with no changes in the forces, then there is an equilibrium probability distribution of the present states, the so-called steady-state distribution, that is reached irrespective of the original state of the population. So, if we can observe many genetic variations all of which can be assumed to be the result of the same forces, then the distribution of those variations can be used to estimate those forces. Finally, we require a set of statistical procedures that can test the agreement between the static observations and various hypotheses about the strength of the different forces, especially whether the observations indicate the operation of natural selection as opposed to purely random drift events.

To drive home this point, let me give you a concrete case study of the difficulty of 'inverse evolutionary inference' in the case of the evolution of language, specifically, the case of FOXP2. I am sure you are all familiar with the basics of the *FOXP2* story. Damage to the *FOXP2* gene disrupts normal language abilities, as first picked up in members of the KE family (though the effects are complex, and still being unraveled). FoxP2 is a transcription factor, a gene 715 amino acids long on human chromosome 7q.31 that is the blueprint for a protein that in turn regulates the activity of other genes by binding to DNA. [SLIDE]. By gripping DNA like a clothing pin, FoxP2 can either up- or down-regulate the activity of other genes (and the proteins they code for).

But let's focus on the evolutionary picture. From a comparative evolutionary standpoint, Enard *et al.* in *Nature,* 2002 showed that there are just two specific *FOXP2* protein-coding differences between us and other non-human primates (chimp, organg, rhesus macaque), at 'amino acid' positions 303 and 325 of a certain portion of the *FOXP2* gene, exon 7. These are places in the DNA where a single 'letter' (nucleotide) change, the raw stuff of evolutionary change, leads to a change in the corresponding amino acid that 3-letter stretch of DNA codes for, a so-called ''replacement' or nonsynonymous' substitution. Here it is the Think of this as changing the 'semantics' of the blueprint. (One of these changes, at position 325 has putative functional consequences but in the original 2002 paper this was inferred by a computer simulation.) Because the DNA code is redundant, note that other DNA letter (nucleotide) changes might not alter the resulting amino acid coded for – the redundancy is a built-in error-correction mechanism, e.g., if the DNA triplet AAT coding for Asparagine has its last letter T(hymine) mutated to C(ytoseine), it still codes for the same amino acid, a *silent* or *synonymous* nucleotide change. Such changes are presumed to be 'neutral' since they don't result in any functional change and so are assumed invisible to the sieving hand of natural selection (note though that even this assumption that these so-called 'neutral' changes are invisible to selection is itself questionable). Be that as it may, ironically, chimpanzees' FOXP2 genes are more like mice than humans. (Also importantly, these two nonhuman to human DNA changes are <u>not</u> the ones that cause the deficits observed in the KE families; the affected KE family individuals' DNA is in this sense wholly human.) Taken together with the rest of the comparative genomic analysis – the lack of variation in the gene within a geographically distributed range of human individuals – and applying some of the standard tools of the trade for detecting the 'signature' of natural selection, Enard *et al*. (2002) advanced the important evolutionary suggestion that there is strong evidence for the operation of natural selection on the gene at these two regions where *FOXP2* differs between us and non-human primates, a <u>selective</u> <u>sweep</u>, that took place somewhere around 0 – 200 thousand years ago.

So this is precisely an example of the sort I just described. [SLIDE]: we use observed differences in living species to infer an evolutionary change that occurred in the past. Just how confident can we be in about this inference? The answer is that these evolutionary conclusions are pretty fragile. Just to see how this time line stacks up against other estimates of the phylogenetic relationship between us and our extinct relatives, on this slide I show the 'point estimate' published by the Chimpanzee genome consortium in 2005. Further recall that the recent sequencing of Neandertal DNA showed that they shared the same two FOXP2 changes as us; Krause *et al* 2007 used this data to get their own estimate of when the 'selective sweep' for FOXP2 occurred. Note that this estimate at 600KYA or more is far more ancient than the 'oldest' time estimate reported by Enard *et al.* The waters would turn muddier still if it turns out that the recently discovered 'Denisova' homin called *X*-woman has FOXP2 DNA that matches our own (and Neandertals), since the estimated divergence between the least common ancestor of us and Neandertals and the *Denisova* sample has been estimated at 1.04 MYA (based on mitochondrial DNA samples). This would seem to call into question the, uhm, sweeping conclusions of the earlier results. (But it would be entirely compatible with the broader view of FOXP2's role in language that I suggest below.)

To understand why this evolutionary inference should turn out to be so fragile, we have to ask: what's a selective sweep? No, stop thinking sports. Here's a cartoon of the particular method used by Enard *et al.* 2002 to figure out the timing of the possible FOXP2 selective sweep [SLIDE]. The 8 lines represent different DNA strands. The blue dot corresponds to the FOXP2 favorable gene variant. A red dot is some other so-called 'neutral' segment of DNA, i.e., neither beneficial or deleterious, and so 'invisible' to selection's sieve. We imagine that at first there are just a few DNA strands with the new FOXP2 variant (allele). If it's strongly beneficial, and selected then it will leave more copies in later generations (the next panel), rapidly rising to fixation in the population. But here's the trick: it might also carry along with it nearby red dots, close enough so that they too will get preferentially copied as well – picturesquely dubbed 'genetic hitch-hiking' (though I prefer Gillespie's term 'genetic draft' – they are ***swept*** along with the flow). So some lucky, otherwise 'neutral' variant in virtue of being closely linked to a real winner, FOXP2, also gets star treatment, and rises in frequency too, becoming differentially skewed, 'more frequent than expected'. Eventually, the beneficial FOXP2 variant propagates to everybody, "fixes" along with the red-dot variant, as shown in third snapshot at the end of the 'selective sweep'. From then on, FOXP2 is maintained in the population as we've seen, but the strong association between the neutral, hitchhiked variant and FOXP2 can slowly decay, due to the inevitable mash-ups caused by the swapping of DNA segments between chromosome pairs (sex rears its head). If the decay time is short enough – say within the last 300 thousand years, then in the here-and-now, we can hope to still find a frequency skew in these 'neutral' variants - a lower-than-usual 'diversity' of 'neutral' DNA segments, and a higher than usual frequency of rare neutral DNA segments, all nearby FOXP2. That's what's actually measured. (Question: So how can this method ever detect a sweep more than 300 KYA? Ie, if we are talking about the point before us and Neandertals diverged, given that they seem to have had the same *FOXP2* DNA as us? Good question. The 'sweep method' cannot, in general, probe back that far.)

Now, in order to estimate the time *when* the sweep ended, we must carry out an 'inverse' simulation by computer – in this case, 3 million of them at time steps 1000, 2000, …, 6000 generations back in time, fixing certain assumptions about the (effective) population size, the number of generations, the strength of selection, and so forth. And all the assumptions matter, possibly because the 'signal' of selection is so weak. Enard *et al* themselves note they get their highest likelihood result with the selection coefficient set to 1%. This is not unreasonable, but it appears to be on the high side when compared to known field-measured values. For example, Satta *et al* 1993 measured the intensity of selection at 7 of the MHC major histocompatability loci, and obtained estimates ranging from 0.07% to 4.2%, more typically, under 1%. And other, typical field-measured values here are an order of magnitude less. When we set the selectional strength to such values as 10^{-3}, the sweep disappears. Similarly, if we dial up the simulation mesh timestep to once every 100 instead of once per 1000 generations, the sweep disappears.

In fact, Molly Przeworski at Chicago, a leader in this field who was a co-author on the Enard *et al.* paper in 2002, who last year pointed out flaws in the Neandertal 'sweep' data characterizes the situation in this way:

"The approach will only be reliable if the signature of natural selection is accurately characterized. Currently, our understanding of the effects of directional positive selection stems from the simple model of a random-mating population of constant size."

Further, she points to a telling example of a 'false negative': "one of the rare cases with independent evidence for a recent selective sweep" – interestingly another case of resistance to malaria, the so-called Duffy blood group antigen, a protein on the surface of red blood cells coded by a gene on chromosome 1. She notes, "it is troubling that in this case the expected signature of a selective sweep was not observed. (Hamblin and Di Rienzo 2000)." M. Przeworski, G. Coop, & J. D. Wall, 2005:2312, "The signature of positive selection on standing genetic variation" *Evolution* 59:1, 2312.

My conclusion is that at the present moment we simply cannot rely on any of the 'bars' in the selection scenarios. It's a situation where what we would like to know outstrips our current signal detection methods. As Richard Lewontin recently remarked in the New York Review of Books, "There are different modes of knowing and we 'know' that evolution has in fact occurred in a stronger sense than we 'know' that some sequence of evolutionary change has been the result of natural selection." I suspect this may be one of those cases. No surprise: we can be far more confident that the human *testes* have been under positive selection than the brain. But this, every woman already knows.

OK, now let's turn to Part II and to the "good news" coming from modern linguistic theory, bearing in mind the special explanatory challenges.

Despite the issues in dating the FOXP2 changes, according to a fairly general consensus among paleoanthropologists and archaeologists, the origin of language is very recent in evolutionary time. As far as is known, the language faculty has remained essentially unchanged since then – which is not surprising in such a brief period. An infant from a stone age tribe in the Amazon, if brought to Boston, will be indistinguishable in linguistic and other cognitive functions from children born in Boston who trace their ancestry to the first English colonists; and conversely. The actual dates are uncertain, and do not matter much for our purposes. The general picture appears to be roughly accurate, though not everyone here would agree about them. Language is a species property of humans, a common endowment with no significant variation apart from serious pathology, unlike anything else known in the organic world in its essentials, and surely central to human life since its emergence. This complex seems to have crystallized fairly recently among a small group, perhaps roughly 10,000-18,000 in effective population size, of whom we are all descendants.

There are, of course, plenty of animal communication systems. But they are all radically different from human language in structure and function. Human language does not even fit within the standard typologies of animal communication systems – Marc Hauser's, for example, in his comprehensive review of the evolution of communication (1997). It has been conventional to regard language as a system whose function is communication. This is indeed the widespread view invoked in most selectionist accounts of language, which almost invariably start from this interpretation. However, to the extent that the characterization has any meaning, this appears to be incorrect, for a variety of reasons to which we turn below.

Moreover, there has always been an alternative tradition, expressed by Burling (1993) among others, that humans may well possess a secondary communication system like those of other primates, namely a nonverbal system of gestures or even calls, but that this is not language, since, as Burling notes, "our surviving primate communication system remains sharply distinct from language."[1]

Language can of course be used for communication, as can any aspect of what we do: style of dress, gesture, and so on. And it can be and commonly is used for much else. Statistically speaking, for whatever that is worth, the overwhelming use of language is internal – for thought.

---

[1] Laura Petitto's work on the acquisition of sign language (1987) demonstrates Burling's point rather dramatically – the same gesture is used for pointing and pronominal reference, but in the latter case the gesture is counter-iconic at the age when infants typically reverse "I" and "you".

It takes an enormous act of will to keep from talking to oneself in every waking moment – and asleep as well, often a considerable annoyance. The distinguished neurologist Harry Jerison (1977:55) among others expressed a stronger view, holding that "language did not evolve as a communication system…. the initial evolution of language is more likely to have been…for the construction of a real world," as a "tool for thought."

How, then, did this strange object appear in the biological record, apparently within a very narrow evolutionary window, perhaps about 50,000–100,000 years ago? There are of course no definite answers, but it is possible to sketch what seem to be some reasonable speculations.

What was the change? To answer that, we have to consider the special properties of language. What do we have that other animals don't? What is the language phenotype? Well, we can do this. [SLIDE of Dan Brown] Or I can show you in one slide, one that Marc Hauser might have shown you [SLIDE of pencil.] We have a combinatorial promiscuity that is unparalleled: we construct not just new sentences, but new tools, out of old ones, endlessly. The most elementary property of our shared language capacity is that it enables us to construct and interpret a discrete infinity of hierarchically structured expressions: discrete because there are 5 word sentences and 6 word sentences, but no 5½ word sentences; infinite because there is no longest sentence. Language is therefore based on a recursive*, generative* procedure that takes elementary word-like elements from some store, and applies repeatedly to yield structured expressions, without bound. To account for the emergence of the language faculty – hence for the existence of at least one language – we have to face two basic tasks. One task is to account for the "atoms of computation," commonly in the range of 30–50,000. I will put to one side for now this important matter, returning to this key question at the end. The second task is to discover the computational properties of the language faculty. This task in turn has several facets: we must seek to discover the generative procedure that constructs infinitely many expressions in the mind, and the methods by which these internal mental objects are related to two *interfaces* with language-external (but organism-internal) systems: the system of thought, on the one hand, and also to the sensory-motor system, thus *externalizing* internal computations and thought. This is one way of reformulating the traditional conception, at least back to Aristotle, that language is sound with a meaning.

Francois Jacob (1977) provided a suggestive model for the development of other organisms based on the notion that "thanks to complex regulatory circuits" what "accounts for the difference between a butterfly and a lion, a chicken and a fly … are the result of mutations which altered the organism's regulatory circuits more than its chemical structure" (1977:26). Jacob's model, presented at a joint conference on biolinguistics in 1975, in turn provided part of the inspiration for the Principles and Parameters approach to language, a matter discussed in lectures shortly after (Chomsky*,* 1980**:**67), and that illustrated by the slide I presented early on, repeated here. [SLIDE]

This approach was based on the assumption that languages consist of fixed and invariant principles connected to a kind of switchbox of parameters, questions that the child has to answer on the basis of presented data in order to fix a language from the limited variety available in principle – or perhaps, as Charles Yang has argued (2002), to determine a probability distribution over languages resulting from a learning procedure for parameter setting. For example, the child has to determine whether the language to which it is exposed is "head-initial," like English, a language in which substantive elements precede their objects, as in "read the book" or "in the room”; or whether it is “head-final,” like Japanese, where the counterparts would be “book read” and “room in.” As in the somewhat analogous case of rearrangement of regulatory mechanisms, the approach suggests a framework for understanding how essential unity might yield the appearance of the limitless diversity that was assumed not long ago for language (as for biological organisms generally).

The research program that arose from this viewpoint has been very fruitful, yielding rich new understanding of a very broad typological range of languages, opening new questions that had never been considered, sometimes providing answers. It is no exaggeration to say that more has been learned about languages in the past twenty-five years than in the earlier millennia of serious inquiry into language. The approach suggests that what emerged was the generative procedure that provides the principles, and that diversity of language results from the fact that the principles do not determine the answers to all questions about language, but leave some questions as open parameters. Note that the single illustration above has to do with ordering. Though the matter is contested, it seems that there is by now substantial linguistic evidence that ordering is restricted to externalization of internal computation to the sensory-motor system, and plays no role in core syntax and semantics, a conclusion for which there is also accumulating biological evidence of a sort familiar to mainstream biologists, to which we return below.

The simplest assumption, hence the one we adopt unless counterevidence appears, is that the generative procedure emerged suddenly, in the evolutionary sense. Various kinds of generative procedures have been explored in the past 50 years. One approach familiar to linguists and computer scientists is phrase structure grammar, developed in the 1950s and since extensively employed. The approach made sense at the time. It fit very naturally into one of the several equivalent formulations of the mathematical theory of recursive procedures – Emil Post’s rewriting systems – and it captured at least some basic properties of language, such as hierarchical structure and embedding. Nevertheless, it was quickly recognized that phrase structure grammar is not only inadequate for language but is also quite a complex procedure with many arbitrary stipulations, not the kind of system we would hope to find.

Over the years, research has found ways to reduce the complexities of these systems, and finally to eliminate them entirely in favor of the simplest possible mode of recursive generation: an operation that takes already-constructed two objects, call them *X* and *Y*, and forms from them a new object that consists of the two unchanged, hence simply the set with *X* and *Y* as members. Call this operation whatever you want; I will call it *cons*, after the ‘*cons*tructor’ operation in the programming language LISP. (Chomsky calls it Merge; but in fact, it goes by many other names, since it appears in virtually every full-fledged, modern linguistic theory. While linguists differ on many things, they are in general agreement that *some* such combinatorial operator is required. So I will use this more ‘neutral’ term instead, as a gloss covering the combinatorial operation that all modern linguistic theories contain, from LFG to HPSG, to combinatory categorial grammar, and the like.)

Provided with conceptual atoms of the lexicon, the operation *cons,* iterated without bound, yields an infinity of hierarchically constructed expressions. Here’s how it works. If we have two items, say a verb ‘ate’ and a Noun Phrase ‘the apple’, then *cons* combines them into a new object that has verb-like properties – with the label ‘VP’ (verb phrase), ‘ate the apple’. Crucially, from then on, the syntax manipulates this new construct as if were a single new object with *just* the label ‘VP’ (alternatively, all the features of the verb), which can then enter into further operations; it never looks back to 'see' the two pieces 'ate' and 'the apple' from which the VP was built. And *that's* the innovation: the ability to form *new* objects with novel 'labels' that are composed of parts, and then use them, in turn, to form yet further novel syntactic objects. Each syntactic object X formed by the repeated application of *cons* has properties that enter into further computation, including semantic/phonetic interpretation: a verb phrase VP functions differently from a noun phrase NP. If these objects can be interpreted by conceptual systems, the operation provides an internal "language of thought."

A very strong thesis, called the "strong minimalist thesis," is that the generative process is optimal in the sense that the principles of language are determined by efficient computation and language keeps to this single, simplest recursive operation, designed so its output representations are 'visible' or 'legible' to the sensory-motor system and some conceptual-reasoning system (roughly, it is pronounceable and thinkable). In short, as far as *linguistically special* operations go, this is *all* you need. Now, the strong minimalist thesis is very far from established, but it looks much more plausible than it did only a few years ago. Insofar as it is correct, the evolution of language will reduce to the emergence of the operation *cons*, the evolution of conceptual atoms of the lexicon, the linkage to conceptual systems, and the mode of externalization. Any residue of principles of language not reducible to *cons* and optimal computation will have to be accounted for by some other evolutionary process.

Note that there is no room in this picture for any precursors to language – say a language-like system with only short sentences. There is no rationale for such a system: to go from seven-word sentences to the discrete infinity of human language requires emergence of the same recursive procedure as to go from zero to infinity, and there is of course no direct evidence for such "protolanguages." Similar observations hold for language acquisition, despite appearances, a matter that we put to the side here.

Crucially, the operation *cons* immediately yields the familiar *displacement* property of language: the fact that we pronounce phrases in one position, but interpret them somewhere else as well. Thus in the sentence "guess what John is eating," we understand "what" to be the object of "eat," as in "John ate the apple," even though it is pronounced somewhere else. This property has always seemed paradoxical, a kind of "imperfection" of language. It is by no means necessary in order to capture semantic facts, but it is ubiquitous. It surpasses the capacity of simple phrase structure grammars, requiring that they be still further complicated with additional devices. But it falls within *cons* automatically.

To see how, suppose that the operation *cons* has constructed the mental expression corresponding to "John ate what." A larger expression can be constructed by Merge in two ways: *Internal cons* can add something from within the expression, so as to form "what John ate what"; and *External cons* can add something new, yielding "guess what John ate what."

That carries us part of the way towards displacement. In "what John ate what," the phrase "what" appears in two positions, and in fact those two positions are required for semantic interpretation: the original position provides the information that "what" is understood to be the direct object of "ate," and the new position, at the edge, is interpreted as a quantifier ranging over a variable, so that the expression means something like "for which thing *x*, John ate the thing *x*."

These observations generalize over a wide range of constructions. The results are just what is needed for proper semantic interpretation. But note that these do not yield the objects that are pronounced in English. We do not pronounce "guess what John ate what," but rather "guess what John ate," with the original position suppressed. That is a universal property of displacement, with minor (and interesting) qualifications that we can ignore here. The property follows from elementary principles of computational efficiency. In fact, it has often been noted that serial motor activity is computationally costly, a matter attested by the sheer quantity of motor cortex devoted to both motor control of the hands and for oro-facial articulatory gestures.

To say something, get it into the outside world or 'externalize' it we must take this hierarchical object inside our heads and flatten it onto the left-to-right system of the motor articulators that are, in most cases, the muscles that move our tongues and vocal tract. The laws of physics dictate that this sound stream proceeds left to right in time. (We can't pile up all the sounds on top of each other, or literally speak out of both sides of our mouth.) To externalize the internally generated expression "what John ate what," it would be necessary to pronounce "what" twice, and that turns out to place a very considerable burden on computation, when we consider expressions of normal complexity and the actual nature of displacement. With all but one of the occurrences of "what" suppressed, the computational burden is greatly eased. The one occurrence that must be pronounced is typically the most prominent one, generally (but not always) the last one created by Internal *cons*: otherwise there will be no indication that the operation has applied to yield the correct interpretation. It appears, then, that the language faculty recruits a general principle of computational efficiency for the process of externalization.

The suppression of all but one of the occurrences of the displaced element is computationally efficient, but imposes a significant burden on interpretation, hence on communication. The person hearing the sentence has to discover the position of the gap where the displaced element is to be interpreted. That is a highly non-trivial problem in general, familiar from parsing programs, as can be illustrate here. [SLIDES] There is, then, a conflict between computational efficiency and interpretive-communicative efficiency. Universally, languages resolve the conflict in favor of computational efficiency. These facts at once suggest that language evolved as an instrument of internal thought, with externalization a secondary process.

There is a great deal of evidence from language design that yields similar conclusions, so called "island properties," for example, indicating that there are tight locality bounds on aspects of displacement to ease the computational burden; evidently, one cannot displace an element "too far". [SLIDE]

There are independent reasons for the conclusion that externalization is a secondary process. One is that externalization appears to be modality-independent, as has been learned from studies of sign language in recent years. The structural properties of sign and spoken language are remarkably similar. Additionally, acquisition follows the same course in both, and neural localization seems to be similar as well. That tends to reinforce the conclusion that language is optimized for the system of thought, with mode of externalization secondary.

Note further that the constraints on externalization holding for the auditory modality also appear to hold in the case of the visual modality in signed languages. Even though there is no physical constraint barring one from 'saying' with one hand that 'John ate the apple' and 'Mary ate the pear' with the other hand, nevertheless it appears that one hand is dominant throughout and delivers sentences (via gestures) in a left-to-right order in time, linearized as in vocal-tract externalization, while the non-dominant hand adds markings for emphasis, morphology, and the like.

It seems possible to make a far stronger statement: all recent relevant biological and evolutionary research leads to the conclusion that the process of externalization is secondary. This includes the *FOXP2* story we talked about earlier. Recent discoveries in birds and mice over the past few years point to an "emerging consensus" that this transcription factor gene is not part of a blueprint for internal syntax, the narrow faculty of language, and most certainly not some hypothetical 'language gene' (just as there no single genes for eye color or autism) but rather part of regulatory machinery related to externalization and vocal learning (Vargha-Khadem, 2005; Groszer, *et al.* 2008). *FOXP2* aids in the development of serial fine motor control, oro-facial or otherwise: the ability to literally put one 'sound' or 'gesture' down in place, one point after another in time.

In this respect it is worth noting that members of the KE family in which this genetic defect was originally isolated exhibit a quite general motor dyspraxia, not localized to simply their oro-facial movements. Recent studies where a mutated *FOXP2* gene built to replicate the defects found in the KE family was inserted in mice confirm this view: "We find that Foxp2-R552H heterozygous mice display subtle but highly significant deficits in learning of rapid motor skills... These data are consistent with proposals that human speech faculties recruit evolutionarily ancient neural circuits involved in motor learning" (Groszer, *et al*., 2008:359).

If this view is on the right track, then *FOXP2* would be more akin to the blueprint that aids in the construction of a properly functioning input-output system for a computer, particularly critical for vocal learners who must match what they hear and practice. But it is not part and parcel of the central faculty of language itself. From this point of view, what has gone wrong in the affected KE family members is thus something awry with the externalization system, not the central language faculty itself. If this is so, then the evolutionary analyses suggesting that this transcription factor was under positive selection approximately 100,000 years ago (in itself arguable) could in fact be quite inconclusive about the evolution of the core components of the faculty of language, syntax and the mapping to the "semantic" (or conceptual-inference) interface. It is challenging to determine the causal sequence: the link between *FOXP2* and high-grade serial motor coordination could be regarded as either an opportunistic pre-requisite substrate for externalization, no matter what the modality, as is common in evolutionary scenarios, or the result of selection pressure for efficient externalization 'solutions' after Merge arose. In either case, *FOXP2* becomes part of a system extrinsic to core syntax/semantics.

In many respects this focus on *FOXP2* and dyspraxia is quite similar to the near-universal focus on 'language as communication'.[2] Both efforts examine properties apparently particular only to

---

[2] This is much like attending solely to the different means by an LCD television and the old cathode-ray tube TVs display moving images without paying any attention to what image is being displayed. The old TVs 'painted' a picture by sweeping an electron beam over a set of chemical dots that would glow or not. Liquid crystal displays operate by an entirely different means: roughly, they pass light or not through a liquid crystal array of dots depending on an electric charge applied to each 'dot,' but there is no single sweeping beam. One generates the same flat image by an entirely different means. Similarly, whether the externalized, linear timing slots are being set out by motor commands to the vocal tract or by moving fingers is irrelevant to the more crucial 'inner' representations.

the externalization process, which, we conjecture, is not part of the core faculty of human language. In this sense both efforts are misdirected, unrevealing of the internal computations of the mind/brain, the cognitive revolution notwithstanding. By expressly stating the distinction between internal syntax and externalization, many new research directions may be opened up, and new concrete, testable predictions posed particularly from a biological perspective, as the example of animal vocal learning illustrates.

Returning to the core principles of language, unbounded *cons* (hence displacement) would give even a single individual so endowed with many advantages: capacities for complex internal thought, planning, interpretation, and so on; there would be no paradox of having 'no one to talk to.' The capacity would be partially transmitted to offspring, and because of the selective advantages it confers, it might come to dominate a small breeding group. This is the scenario that Számadó and Szathmáry (2006) envision that regards language as an 'internal mental tool.'

When the beneficial mutation had spread through the group, there would be an advantage to externalization, so the capacity would be linked as a secondary process to the sensory motor system for externalization and interaction, including communication as a special case. 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.

Investigation of language design can yield evidence on the relation of language to the sensory motor system and thought systems. I think there is mounting evidence to support the natural conclusion that the relation is asymmetrical in the manner illustrated in the critical case of displacement.

Externalization is not a simple task. It has to relate two quite distinct systems: one is a sensorimotor system that appears to have been basically intact for hundreds of thousands of years; the second is a newly emerged computational system for thought, which is perfect, insofar as the strong minimalist thesis is correct. We would expect, then, that morphology and phonology – the linguistic processes that convert internal syntactic objects to the entities accessible to the sensorimotor system – might turn out to be quite intricate, varied, and subject to accidental historical events. Parameterization and diversity, then, would be mostly – possibly entirely – restricted to externalization. That is pretty much what we seem to find: a computational system efficiently generating expressions interpretable at the semantic/pragmatic interface, with diversity resulting from complex and highly varied modes of externalization, which, furthermore, are readily susceptible to historical change.[3]

If this picture is more or less accurate, we may have an answer to the second of the two basic questions: Why are there so many languages? The reason might be that the problem of externalization can be solved in many different and independent ways, either before or after the dispersal of the original population.

---

[3] Positing an independent, recursive "language of thought," as a means to account for recursion in syntax leads to an explanatory regress as well as being unnecessary and quite obscure.

We have no reason to suppose that solving the externalization problem involved an evolutionary change – that is, genomic change. It might simply be a problem addressed by existing cognitive processes, in different ways, and at different times. There is sometimes an unfortunate tendency to confuse literal evolutionary (genomic) change with historical change, two entirely distinct phenomena. As already noted, there is very strong evidence that there has been no relevant recent evolution of the language faculty, though undoubtedly there has been a great deal of change, even invention of modes of externalization (as in sign language). Confusion about these matters could be overcome by replacing the metaphoric notions "evolution of language" and "language change" by their more exact counterparts: evolution of the organisms that use language, and change in the ways they do so. In these more accurate terms, emergence of the language faculty involved evolution, while historical change (which continues constantly) does not.

Again, these seem to be the simplest assumptions, and there is no known reason to reject them. If they are generally on the right track, it follows that externalization may not have evolved at all; rather, it might have been a process of problem solving using existing cognitive capacities. Evolution in the biological sense of the term would then be restricted to the mutations that yielded the operation *cons*, along with whatever residue resists explanation in terms of the strong minimalist thesis and any language-specific constraints that might exist on the solution to the cognitive problem of externalization. Accordingly, any approach to "evolution of language" that focuses on communication, or the sensory motor system, or statistical properties of spoken language, and the like, may well be seriously misguided. That judgment covers quite a broad range, as those familiar with the literature will be aware.

But where on earth did *cons* come from? If we break *cons* down into its constituent parts, we might gain some clues that it, too, is built out of parts that were already present in other animals. Let's see how far we can carry this line of thinking. The evo-devo story has provided compelling evidence for two relevant conclusions. One is that genetic endowment even for regulatory systems is deeply conserved. A second conclusion is that very slight changes can yield great differences in observed outcome – though phenotypic variation is nonetheless limited, in virtue of the deep conservation of genetic systems. To cite a well-known example, it turns out that there are very few types of eyes, in part because of constraints imposed by the physics of light, in part because only one category of proteins, opsin molecules, can perform the necessary functions. The genes encoding opsin had very early origins, and are repeatedly recruited, but only in limited ways, again because of physical constraints. The same is true of eye lens proteins. The evolution of eyes illustrates the complex interactions of physical law, stochastic processes, and the role of selection in choosing within a narrow menu of possibilities (Gehring, 2005). (Gehring argues that the original novelty itself – a photoreceptor next to a pigment cell – was probably a stochastic event, and further, there is reasonable evidence from genome analysis that it might in part be due to symbiosis – the wholesale incorporation of many genes into an animal (eukaryotic) cell by ingestion of a chloroplast from *Volvox*, a cyanobacter.)

So let's see if we can decompose *cons* into its atomic parts and speculate as to whether these pieces are found in other animal species. Now, *cons* does 3 things with the objects it glues together: it groups then, selects one of them as more prominent, and third, labels the new glued-together syntactic object with a new name, usually that of the selected object. But we can immediately see that grouping and selection are found everywhere besides language, and in other animals. As for us, we can look at poetry and analyze them along the lines suggested by Morris

Halle. We start with a string of consonant-vowel pairs, without any other features, so I'll just mark them with asterisks as Morris does.

*   *   *   *   *   *

Now these syllables are processed according to a very simple set of rules. To take this example, going left to right, we group two asterisks at a time, selecting one as 'most prominent', then promote this selected asterisk to the next line. We repeat the process until we can carry out no more grouping operations. After the dust settles, we have formed a layer of 'metrical tiers'. (Different choices of selection and grouping, left-to-right, or selecting the rightmost element as most prominent give rise to different metrical patterns. Halle gives many good arguments showing that this determines the entire range of possible metrical structures in all languages. It is important to realize that as far as metrical verse goes, this is a machine for grouping syllables, not an abstract model of the concrete stress patterns of words, though it might be part of such a process. Of course this is *exactly* the same formal apparatus that we saw before in the realm of syntax: an operation that groups two elements, selects one, and repeats. There's one crucial difference: in this system *there are no words* and *no word features*, just 'marks' (the asterisks), so the system is quite limited compared to the one wired to a lexicon. For example, there can be no 'recursion', since that requires one to come up with new labels, bundles of word features, for the glued-together objects to act as new, single atomic units – and here there are no new labels, just the asterisks. Similarly, there is no 'displacement' (movement) and no semantics So this system us not full-fledged language, but rather like a song without words. Now, if we wire the same group-plus-select machinery to another system that lacks words – this one – [SLIDE] we get musical patterns. And yet another: wired to this system of motor action [SLIDE] we get the metrical pattern of gestured language. [SLIDE on stress patterns and the Darth Vader/Chewbacca solution?]

Indeed, this same ability at grouping-plus-selection seems to appear in other species, and so may have been antecedently available to us as well. Take this species of vocal learners, songbirds. Here is a frequency analysis of a zebra finch song, with the bottom portion marking out the 'power peaks', evidently also with metrical structure, though it again remains important to disentangle the metrical structure from its 'performance' in terms of stress, either produced or perceived. [SLIDE]. Putting this question to one side, one might again note that since birds have no 'words', no 'lexicon', and no 'word features', that their songs too would not amount to full-fledged language, lacking semantics or displacement – songs without words. But they carry us nearly all the way to language.

One crucial missing link remains: the storehouse of words, the 'lexical atoms' that *cons* uses as its labels. On the story presented here, this is the innovation, that, when added to the pre-existing substrate of grouping+selection, yields the recursive hierarchical structure that characterizes human language. So where do words come from?

Conceptual structures are found in other primates: probably actor-action-goal schemata, categorization, possibly the singular-plural distinction, and others. These were presumably recruited for language**,** though the conceptual resources of humans that enter into language use appear to be far richer. Specifically, even the "atoms" of computation, lexical items/concepts, appear to be uniquely human. And I think that leaves us with this [SLIDE OF APPLE] – something like a version of original sin. Here I could do no better than to quote from what Prof. Laura Petitto has written: "Although chimps can learn to use the sign *apple* while in front of a red or a green apple, children learn this effortlessly…Chimps, unlike humans, use such labels in a way that seems to rely heavily on some global notion of *association*. A chimp will use the same label *apple* to 'mean' the action of eating apples, the location where apples are kept, events and locations of objects other than apples that happen to be stored with an apple (like the knife used to cut it) and so on, all simultaneously, without apparent recognition of the relevant differences…surprisingly, then, chimps do not really learn the human word *apple* at all." (1997, 86). So perhaps it is really true that this [SLIDE] *is* the original sin leading to language: a bite of the apple.

Crucially, even the simplest words and concepts of human language and thought lack the relation to mind-independent entities that appears to be characteristic of animal communication. The latter is held to be 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 cognitive neuroscientist Randy Gallistel, introducing a major collection of papers on animal communication (Gallistel, 1990). According to Jane Goodall, the closest observer of chimpanzees in the wild, for them "the production of a sound in the *absence* of the appropriate emotional state seems to be an almost impossible task" (Goodall, cited in Tattersall, 2002).

The symbols of human language and thought are sharply different. Their use is not automatically keyed to emotional states, and they do not pick out mind-independent objects or events in the external world. For human language and thought, it seems, there is no *reference* relation in the sense of Frege, Peirce, Tarski, Quine, and contemporary philosophy of language and mind. What we understand to be a river, a person, a tree, water, and so on, consistently turns out to be a creation of what 17th century investigators called the human "cognoscitive powers," which provide us with rich means to refer to the outside world from intricate perspectives. As the influential neo-Platonist Ralph Cudworth put the matter, it is only by means of the "inward ideas" produced by its "innate cognoscitive power" that the mind is able to "know and understand all external individual things," articulating ideas that influenced Kant. The objects of thought constructed by the cognoscitive powers cannot be reduced to a "peculiar nature belonging" to the thing we are talking about, as David 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 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 physicist 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 shared cognoscitive powers, and succeeds insofar as shared mental constructs, background, concerns, presuppositions, and so on, allow for common perspectives to be (more or less) attained. These properties of lexical items seem to be unique to human language and thought, and have to be accounted for somehow in the study of their evolution. How, no one has any idea. The fact that there even is a problem has barely been recognized, as a result of the powerful grip of the doctrines of referentialism.

Human cognoscitive powers provide us with a world of experience, different from the world of experience of other animals. Being reflective creatures, thanks to the emergence of the human capacity, humans try to make some sense of experience. These efforts are called myth, or religion, or magic, or philosophy, or in modern English usage, science. For science, the concept of reference in the technical sense is a normative ideal: we hope that the invented concepts *photon* or *verb phrase* pick out some real thing in the world. And of course the concept of reference is just fine for the context for which it was invented in modern logic: formal systems, in which the relation of *reference* is stipulated, holding for example between numerals and numbers. But human language and thought do not seem to work that way, and endless confusion has resulted from failure to recognize that fact.

We enter here into large and extremely interesting topics that will have to put aside. Let us just summarize briefly what seems to be the current best guess about unity and diversity of language and thought. In some largely unknown way, our ancestors developed human concepts. This was cobbled together with a pre-existing system for grouping and selection, and this provided the labels for the operation *cons* – an operation that takes human concepts as computational atoms, and yields structured expressions that provide a rich language of thought. The innovation had obvious advantages, and took over the small group. At some later stage, the internal language of thought was connected to the sensory motor system, a complex task that can be solved in many different ways and at different times, and quite possibly a task that involves no evolution at all. In the course of these events, the human capacity took shape, yielding a good part of our "moral and intellectual nature," in Wallace's phrase (1871). Or, to sum it up much more simply all in one slide, this is what it took for human language to evolve, evolution's own infinite re-use of finite means: [SLIDE]

July, 2008
Room 32D-728 and Room 32D-840
Cambridge Massachusetts, 02139

References

Ahouse, J. and Berwick, R., 1998. Darwin on the mind, *Boston Review of Books*, April/May.

Burling, R., 1993. Primate calls, human language, and nonverbal communication. *Current Anthropology,* 34:1, 25–53.

Carroll, S., 2005. *Endless Forms Most Beautiful*. New York: W.W. Norton & Company.

Chomsky, N., 1980. *Rules and Representations*. New York: Columbia University Press.

Colosimo, P. F., Peichel, C. L., Nereng, K., Blackman, B. K., Shapiro, M. D., Schluter, D., and Kingsley, D. M., 2004. The genetic architecture of parallel armor plate reduction in threespine sticklebacks, *PloS Biology,* 2, 635–641.

Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villareal J. G., Dickson, M., Grimwood, J., Schmutz, J., Myers, R. M., Schluter, D., and Kingsley, D. M., 2005. Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles, *Science* 307, 928–1933.

Darwin, C., 1859. *On the Origin of Species*. London: John Murray.

Darwin, C., 1868. *Variation of Plants and Animals under Domestication*. London: John Murray.

Enard, W., Przeworski, M., Fisher, S.E., Lai, C., Wiebe, V., Kitano, T., Monaco, A., and Paääbo, S., 2002. Molecular evolution of *FOXP2*, a gene involved in speech and language. *Nature*, 418, 869–872.

Fisher, S. E., Vargha-Khadem, F., Watkins, K. E., Monaco, A. P., and Pembrey, M. E., 1998. Localisation of a gene implicated in a severe speech and language disorder. *Nature Genetics*, 18, 168–170.

Gallistel, C. G., 1990. Introduction. In Gallistel, (ed.,) Animal Communication, *Cognition* 37:1-2.

Gehring, W., 2005. New perspectives on eye development and the evolution of eyes and Photoreceptors, *Journal of Heredity*, 96:3, 171–184.

Gillespie, J., 1991. *Population Genetics: A Concise Guide*. 2nd edition. Baltimore: Johns Hopkins University Press.

Groszer, M., Keays, D., Deacon, R., de Bono, J., Prasad-Mulcare, S., Gaub, S., Baum, M., French, D., Nicod, J., Coventry, J., Enard, W., Fray, M., Brown, S., Nolan, P., Pääbo, S., Channon, K., Costa, R., Eilers, J., Ehret, G., Rawlins, N., and Fisher, S.E., 2008. Impaired synaptic plasticity and motor learning in mice with a point mutation implicated in human speech deficits, *Current Biology,* 18, 354–362.

Haldane J. B. S. 1927. A mathematical theory of natural and artificial selection. Part V. Selection and mutation. *Proceedings of the Cambridge Philosophical Society*, 23, 838–844.

Harris, Z., 1951. *Methods in Structural Linguistics*. Chicago: University of Chicago Press.

Hauser, M., 1997. *The Evolution of Communication*. Cambridge, MA: MIT Press.

Hinzen, W., 2006. *Mind Design and Minimal Syntax* Oxford: Oxford University Press.

Jacob, F., 1977. Darwinism reconsidered. *Le Monde*, September, 1977, 6–8.

Jacob, F., 1982. *The Possible and the Actual*. New York: Pantheon.

Jacob, F., 1988. *The Statue Within*. NY: Basic Books.

Jerison, H. J., 1973. *Evolution of the Brain and Human Intelligence*. New York: Academic Press.

Joos, M., 1957. *Readings in Linguistics*, (ed.) Washington: American Council of Learned Societies.

Krause, J., Lalueza-Fox, C., Orlando, L., Enard, W., Green, R., Burbano, H., Hublin, J-J., Hänni, C., Fortea, J., Rasilla, M., Bertranpetit, J., Rosas, A., and Pääbo, S., 2007. The derived FOXP2 variant of modern humans was shared with Neandertals. *Current Biology,* 17, 1–5, November 6.

Lashley, K., 1951. The problem of serial order in behavior, In L.A. Jeffress (ed.), *Cerebral Mechanisms in Behavior*. New York: Wiley.

Lewontin, R., 1998. The evolution of cognition: questions we will never answer. In Don Scarborough Liberman, (eds). *Methods, Models and Conceptual Issues: An Invitation to Cognitive Science*, 4th edition, vol. 4, 108–132.

Lewontin, R., 2001. *The Triple Helix*. New York: New York Review of Books Press.

Luria, S., 1974. *A debate on bio-linguistics*. Endicott House, Dedham Mass, May 20-21, Paris: Centre Royaumont Pour une Science de l'homme.

Lynch, M., 2007. *The Origins of Genome Architecture*. Sunderland, MA: Sinauer Associates.

Maynard-Smith, J., Burian, J. R., Kauffman, S., Alberch, P. , Campbell, J., Goodwin, B., Lande, R., Raup, D., and Wolpert, L., 1985. Developmental constraints and evolution: a perspective from the Mountain Lake Conference on development and evolution. *The Quarterly Review of Biology*, 60:3, 265-287.

Monod, J., 1970. *Le Hasard et La Nécessité*. Paris: Seuil.

Monod, J., 1972. *Chance and Necessity: An Essay on the Natural Philosophy of Modern Biology.* New York: Vintage Books.

Müller, G., 2007. Evo-devo: extending the evolutionary synthesis. *Nature Review Genetics*, 8, 943–949.

Orr, H., 2005. A revolution in the field of evolution? *The New Yorker*, October 24.

Petitto, L. A., 1987. On the autonomy of language and gesture: Evidence from the acquisition of personal pronouns in American Sign Language, *Cognition*, 27:1, 1–52.

Poelwijk, F., Kiviet, D. J., Weinreich, D. M. & Tans, S. J., 2007. Empirical fitness landscapes reveal accessible evolutionary paths. *Nature*, 445:25, 383–386.

Sherman, M., 2007. Universal genome in the origin of metazoa: thoughts about evolution, *Cell Cycle*, 6:15, 1873–1877.

Sauerland, U. and Gärtner, H-M., 2007. *Interfaces + Recursion = Language?* New York: Mouton.

Stevens, K., 1972. The quantal nature of speech: evidence from articulatory-acoustic data, In E.E. David, Jr., and Denes, P.B., eds., *Human Communication: a Unified View*. New York: McGraw-Hill, 51–66.

Stevens, K.N. 1989. On the quantal nature of speech, *Journal of Phonetics* 17:1/2, 3–45.

Striedter, G., 2004. *Principles of Brain Evolution*. Sunderland MA: Sinauer Associates.

Számadó, S. and Szathmáry, E., 2006. Selective scenarios for the emergence of natural language. *Trends in Ecology and Evolution*, 21:10, 555–561.

Tattersall, I., 1998. *The Origin of the Human Capacity* (68th James Arthur Lecture on the Evolution of the Human Brain), American Museum of Natural History.

Tattersall, I., 2002. *The Monkey in the Mirror*, New York: Harcourt.

Tattersall, I., 2005. Patterns of innovation in human evolution, *Evolution und Menschenwerdung*, *Nova Acta Leopoldinia* 345:93.

Thompson, D., 1917/1942. *On Growth and Form.* Cambridge: Cambridge University Press.

Trubetzkoy, N., 1939. Grundzüge der Phonologie. Göttingen: Vandenhoeck & Ruprecht.

Trubetzkoy, N., 1969. *Principles of Phonology*, trans. C. Baltaxe. Berkeley, CA: University of California Press.

Turing, A. and Wardlaw, C. W. (1953/1992). A diffusion reaction theory of morphogenesis, *The Collected Works of Alan Turing: Morphogenesis*. Amsterdam: North-Holland.

Vargha-Khadem, F., Gadian, D.G., Copp, A., and Mishkin, M., 2005. FOXP2 and the neuroanatomy of speech and language. *Nature Reviews in Neuroscience*. 6, 131–138.

Wallace, A. R., 1871. Commentary on *The Descent of Man and Selection in Relation to Sex,* by Charles Darwin, 1871. *The Academy*, 2, no. 20 (15 March), 177–183.

Weinreich, D. M., Delaney, N. F., DePristo, M. A., and Hartl, D. L., 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science*, 7:312, 111-114.

Yang, C., 2002. *Knowledge and Learning in Natural Language*. New York: Oxford University Press.