20

All you Need is Merge: Biology, Computation, and Language from the Bottom-up

ROBERT C. BERWICK

Overture

In recent years, there has been a resurgence of activity surrounding biolinguistics along with a parallel, renewed interest in the connections between language and evolution. To be sure, from one standpoint it has often been said, quite correctly, that linguistic science just *is* biology: the study, however abstract, of a particular, apparently species-specific human trait: knowledge of language. But beyond this immediate point, how might linguistics and biology, especially evolutionary biology, inform one another? How does one go about making a proper evolutionary argument? What can genomics and evolutionary biology tell us now about language, and what might be out of reach, now, or out of reach forever, and why? To answer such questions, this chapter attempts to clear up some possible misunderstandings about evolutionary thinking that one might dub "vulgar Darwinism"—that is, the popular versions of evolutionary theory that sometimes find their way into analyses about language and evolution. The bottom line is that proper evolutionary explanations are often much more difficult to execute than one might think, and that language is a particularly difficult, even uniquely difficult, case to crack. Like linguistics, there is a substantial body of knowledge and theory grounding modern evolutionary analysis, with subtleties that are often missed, even by biologists themselves.

I would like to thank Noam Chomsky, Guglielmo Cinque, Anna Maria Di Sciullo, Morris Halle, Richie Kayne, Andrea Moro, Massimo Piattelli-Palmarini, and Juan Uriagereka for many helpful discussions that sharpened the thoughts in this chapter. Anna Maria Di Sciullo deserves special thanks for fostering and organizing a conference of science and beauty that made such work possible. Any remaining flaws remain the evolutionary deficits of the author, sometimes known as "genetic load."

For example, much excitement has followed from the full genome sequencing of our nearest living relative, the chimpanzee, with other primate genomes to come. However, the special problem of evolutionary inference given close but sparsely populated neighboring species suggests this may tell us very little about human cognitive faculties such as language. The well-known example of a putative "language gene," *FOXP2*, is a prime example: as we shall show, if we re-examine the data from Enard et al. (2002) more carefully, the differences between us and chimps, or for that matter, the more recent similarity between us and Neandertals (Krause, Lalueza-Fox et al. 2007) could be due to chance alone.[1] Where then can we look for insight? The most recent research by Halle in language metrical systems combined with Chomsky's most recent model of syntax may provide a possible and so far unexplored connection to birdsong.

As this chapter immodestly dons the mantle of an opera, it divides into two Acts. Act I opens by outlining the requirements of evolutionary explanations in biology generally and the special problems faced by evolutionary explanations of human language in particular. As we shall see, if one had to choose some trait for evolutionary study, one would be hard pressed to find a trait more challenging than language. Evolution by natural selection can be a potent force, but it is a weak dynamical one, acting over long time frames and easily confounded with demographic effects such as rapid population growth. It therefore becomes difficult to draw inferences about forces in the past given only observations about the present, particularly in the case of humans, who by all accounts passed through just such a population "bottleneck," expanding from a base estimated at 4,000–10,000 individuals and then growing exponentially just at the presumptive time when language emerged and we started on the long trek out of Africa. Moreover, the task becomes even more challenging when the "data points" upon which comparative evolutionary analysis builds are sparse—too few neighboring species compared to other vocal learners like birds. To understand these challenges, Act I lays out the bare bones of evolutionary dynamical theory—the "auto mechanics" required to understand the inferential issues of evolutionary analysis.

Act I concludes with an application of this evolutionary auto mechanics to the recent "banner case" for the genetic, evolutionary analysis of human language, the *FOXP2* transcription factor gene. *FOXP2* has gained much currency in recent years as a putative genomic component that assists in the

[1] We follow convention here by writing the names for human genes in *ITALIC UPPERCASE*, with their corresponding protein products in plain uppercase as in *FOXP2* and FOXP2, with the corresponding non-human genes and protein products written with just initial capitals as *FoxP2* and FoxP2.

construction of the language faculty, if not the language gene itself, and so has been used as a probe for the genomic dissimilarity between us and primates; for detecting the signal of natural selection and perhaps establishing the "starting point" of human language; and even, after extraordinary technical effort, for the evolutionary comparison between modern humans and Neandertals (Fisher et al. 1998; Enard et al. 2002; Krause, Lalueza-Fox et al. 2007). However, on re-analyzing the original Enard et al. (2002) data in light of the extremely small differences that were found between *Homo sapiens* and *Mus musculus* (us and mice)—just two DNA letters, two nucleotides changing just two amino acids—taken together with the special difficulties of evolutionary analysis, Act I concludes that we cannot confidently say that this is a gene 'selected for' language, or even that it was selected for at all. Nor can one say with much confidence that there was a selective sweep that drove this gene to fixation or when it occurred. The differences may well be due to chance alone.

If this is correct, then how does *FOXP2* fit into the picture of language evolution? Act II returns to *FOXP2*, by way of songbirds, to see whether we can face the gap between internal syntax and the external stream of words, the sensor motor interface. Berwick and Chomsky (in this volume) argue that there are many reasons to suppose that *FoxP2* operates quite generally in vocalizing species as part of a system for building an externalization procedure, that is, as part of the sensori-motor system mapping between syntax proper (hierarchical structure generated by Merge) and the output vocal tract or manual gesture articulations. Externalization flattens the hierarchical structures generated by internal syntax by projecting them into a temporally linear succession of articulatory commands (words can only come out one at a time, left-to-right, as it were). Further, Di Sciullo (this volume) already notes that Merge operates at the level of morphology. Taken together, these ideas suggest that one might unpack the syntax–sensori-motor interface even further, into successive stages (perhaps operating in parallel): first moving from syntax, where left-to-right precedence is not expressed; then to morphology, where precedence is expressed. Going one step further, we note that the precedence relation is itself unordered; without further stipulation, a precedence relation does not state whether it is left-to-right or right-to-left. Thus, some additional step of externalization imposes this constraint in order to reach some final, viable, string of temporally ordered motor commands.

In particular, we illustrate this layered analysis of the sensori-motor interface by analyzing metrical stress along the lines of Halle (1997). We show that the Halle system operates very much like Merge, but with one key twist: there

978–0–19–955327–3 20-Sciullo-c20 Di-Sciullo (Typeset by SPi, Chennai) 464 of 491 May 11, 2010 13:25
OUP UNCORRECTED PROOF – FIRST PROOF, 11/5/2010, SPi

can be, obviously, no lexical features, just marks denoted by asterisks, corresponding roughly to consonant–vowel pairs. Following the Halle approach, by a successive sequence of merge operations (projection and selection of heads, as in syntax), we arrive at all and only the possible natural metrical patterns. This is, in effect, pure syntax, with no lexical features and no associated semantics; without features, there is no possibility of internal merge and the movement that we see in ordinary syntax.

We can now raise a key comparative evolutionary question: where else in the biological world might one find metrical structure, but without any lexical information? Act II's proposed answer is: songbirds. Songbirds too produce metrical patterns, but necessarily without lexical information or semantics. Act II then suggests, quite speculatively, that here is precisely where a connection can be made between language, or more precisely the externalization process of language, and FoxP2: recent studies have shown that songbirds (finches) too have altered *FoxP2* DNA sequences (though importantly *no*t the same DNA letters or nucleotides as in humans) and that *FoxP2* disruption also disturbs song learning and its patterning (Teramitsu et al. 2004; Haesler et al. 2007; Vernes et al. 2007). Then, by analyzing birdsong in a novel way, following Coen (2006) to extract its 'songemes,' we can suggest that birdsong too has a metrical structure—in fact exactly that described by the Halle theory as applied to human language. On this view, songbird metrical structure may give us the right kind of comparative, evolutionary insight into at least the externalization process associated with language. Whether this is enough to tell us about Merge itself remains an open question.

20.1 Act I: The Incredible Lightness of Being an Evolutionary Argument

20.1.1 *The challenge for evolutionary explanations and the origin of human language*

Nowhere is the evolutionary explanatory challenge more pointed than in the case of human language. It is practically the polar opposite of a straightforward case such as sickle-cell anemia, where just a single DNA nucleotide change, a DNA letter (adenine to thymine, A to T), leads to a corresponding amino-acid change in the hemoglobin molecule (glutamic acid to valine, this normal amino acid being hydrophobic and thus twisting away from water, the second hydrophilic and thus attracted to water). This single change bends the hemoglobin molecule, in turn visibly crimping the red blood cells—a visibly damaged phenotype or form that shows.

978–0–19–955327–3 20-Sciullo-c20 Di-Sciullo (Typeset by SPi, Chennai) 465 of 491 May 11, 2010 13:25
OUP UNCORRECTED PROOF – FIRST PROOF, 11/5/2010, SPi

But what is the phenotype, the form that shows, in the case of language? Experts cannot even come to an agreement on this most basic of evolutionary questions, whose answer would be the standard starting point for further analysis. Indeed, as noted in the most recent survey article in *Nature* (2008), the link between complex behaviors such as psychiatric conditions and underlying genetics is poorly understood in general. As Harvard neurobiologist and geneticist Steve Hyman remarks, "We are just too ignorant of the underlying neurobiology to make guesses about candidate genes" (Abbott 2008: 157). So how to proceed? Berwick and Chomsky (in this volume) observe that one conventional assumption, "language as communication," leads to difficulties. Alternatively, they adopt the position of Hauser, Chomsky, and Fitch (2002), and construe language as the faculty of language in the narrow sense (FLN), in effect, recursive syntax. In this chapter we shall for the most part simply adopt the FLN view, putting to one side many legitimate questions about this assumption, focusing instead on the evolutionary problematic that remains.

We begin the discussion with language evolution re-described as hypotheses about characters and character states (made explicit with cladograms, that is, branching diagrams of which species held traits in common, and when certain novel traits appeared), but we immediately run into a problem: we have no close living relatives among species, so comparative analysis, the mainstay of the evolutionary program, becomes extremely difficult. Using this representation and the terminology of cladistics invites a discussion of when selectionism provides good explanations for observed traits, in this case, the FLN, and when selectionism is limited.

To explain differences in traits accounted for by adaptation (or natural selection), typically in evolutionary theory we choose exemplars that are as close as possible; to explain adaptationist similarities we choose exemplars that are as far apart as possible. Why is this so? In the best case, what an evolutionary biologist would like to find are examples of what is called "convergent evolution"—multiple, independent solutions to the same functional biological problem. The classic case is that of two forelimbs evolving into wings, as in bats and birds. From this fact we can deduce that four limbs were the ancestral, basal vertebrate state, because both mice, known relatives of bats, and crocodiles, relatives of birds, have four limbs; indeed so do many other vertebrates, providing us with the 'cladogram' depicted in Figure 20.1. Logically this picture says that since all the ancestors of bats and birds had four ordinary limbs, then the two front limbs of bats and birds must have evolved into wings independently, for functional, that is, adaptive reasons. A glance at the similarity in aerodynamic shape of the limbs also calls attention to this adaptive argument, with form following function.

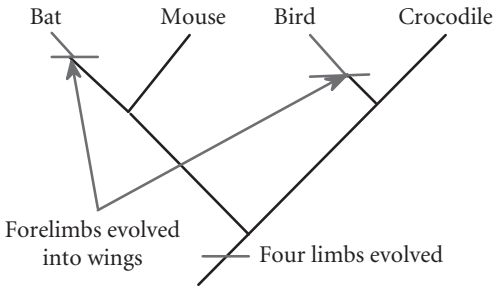


FIGURE 20.1 Multiple independent inventions of wing development from forelimbs established by a cladogram

These two occurrences of wing evolution are therefore multiple *independent* apomorphies. Like independent natural experiments, they offer evidence of convergent evolution where the same evolutionary problem was solved in the same way at different times, powerful evidence that similar functional constraints were at work.[2] In contrast, consider the possible evolutionary explanations for why people and crocodiles share the same number of limbs, four. We might say that four limbs were selected for in both people and crocodiles, for the usual Darwinian reasons. But this would not be a sound explanation. A more satisfying answer is that both species have four limbs not because of some convergent selectional pressure, but because their parents, grandparents, and so forth all had four limbs, back to the common tetrapod ancestor of both. The four-limb solution was not "invented" independently by humans and crocodiles; it is true simply in virtue of shared common descent. Indeed, running the timeline backwards, the reason why we are all tetrapods seems to be a contingent fact about survival during the so-called Cambrian explosion: there were other species with other limb numbers, but all of these five- and seven-limbed alternative species went extinct, for reasons still unclear, but seemingly contingent (Gould 1990).

Returning to comparative evolutionary analysis and language, the point is that it is far easier to run an adaptationist (or selectionist) argument for a trait like wings if one knows, first, whether that trait was simply passed on from a common ancestor or not; and second, whether such a trait has ever

[2] Another example is that of Old World vipers vs. New World pit vipers (*Viperinae* vs. *Crotalinae*). Their last common ancestors are quite ancient (hence the Old World vs. New World split), so we can conveniently use the geographic division to note that the New World vipers must have snake relatives with whom they share common ancestors that are not held in common with the Old World vipers. As with bats and birds, it seems that there are both New World and Old World vipers that have developed "eye lashes" and an arboreal habit. So this must again be an independent development in both lines, just like wings in bats and birds.

evolved independently in remote lineages—independently in the sense that the two species could not possibly have shared the trait in question. Turning now to language, we can immediately see the difficulties. The problem with relatives and near neighbors is that it can be too difficult tell about differences by looking at neighbors—they are too similar and it can be impossible to tell whether a trait is common simply due to common ancestry. This seems to be the case with us, and, say, chimpanzees. In contrast, looking at a very close neighbor without a certain trait usually tells us nothing. From this point of view, a focus on whether chimps, apes, or cotton-top tamarins have some competence similar to us is doomed from the start. Further, if language, or rather the FLN, is indeed a trait unique to the human lineage, a unique, independent autapomorphy, we cannot as easily unfurl the bat wing/bird wing selectionist argument, at least not in the same straightforward way. And there seem to have been no independent inventions of language in the FLN sense of Hauser, Chomsky, and Fitch (2002)—that is, syntax with a recursive operator, which Chomsky calls Merge.

Figures 20.2 through 20.4 depict this explanatory challenge graphically, illustrating the relative sparsity of primate species and the relevant trait of

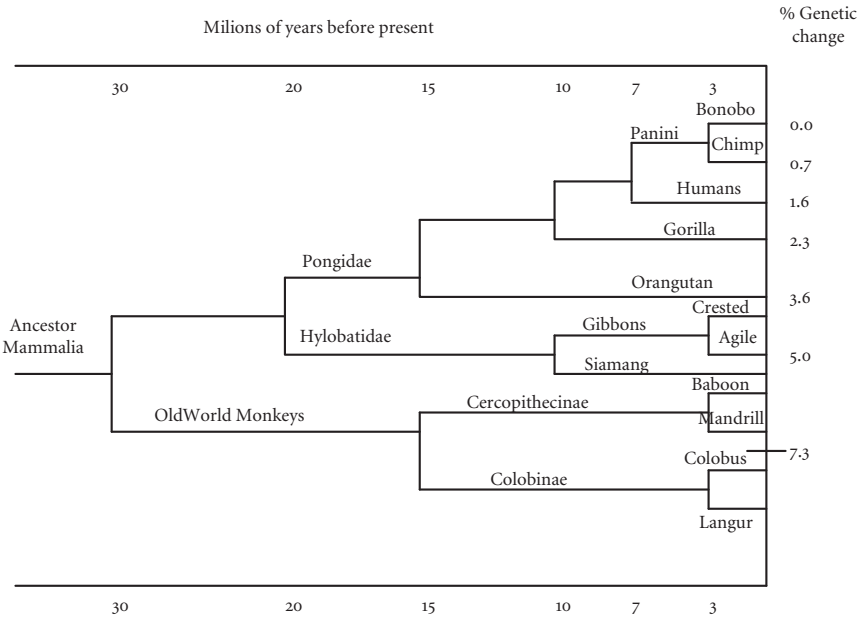


FIGURE 20.2 A phylogenetic diagram of the primate lineage showing rough percentage genetic change among extant groups

vocal learning ability as compared to the same situation in birds. Figure 20.2 exhibits a conventional phylogenetic diagram of the rough genetic distance among living primate lineages. Figure 20.2's horizontal axis runs time backwards from the present, right to left, while the vertical axis measures the percentage genetic difference. (We return to focus on the chimpanzee–human differences in a later section.) Which of these species are true vocal learners, apart from humans, remains unclear (*cf.* the recent reports on gibbons and gibbon syntax by Clarke, Reichard, and Zuberbühler 2006; and the apparent lack of the FLN in cotton-top tamarins, as described in Fitch and Hauser 2004). Whatever the outcome of this ongoing research, the point is that there are only a handful of primate vocal learners/nonlearners to compare, and so, as Lewontin (1998) observes, "it is difficult to connect the dots sensibly."

We can contrast this phylogenetic situation to that in birds, partly shown in Figure 20.3, as redrawn from Zhang, Webb, and Podlaha (2002) and Jarvis (2004), which displays the much richer patterning of vocal learners and

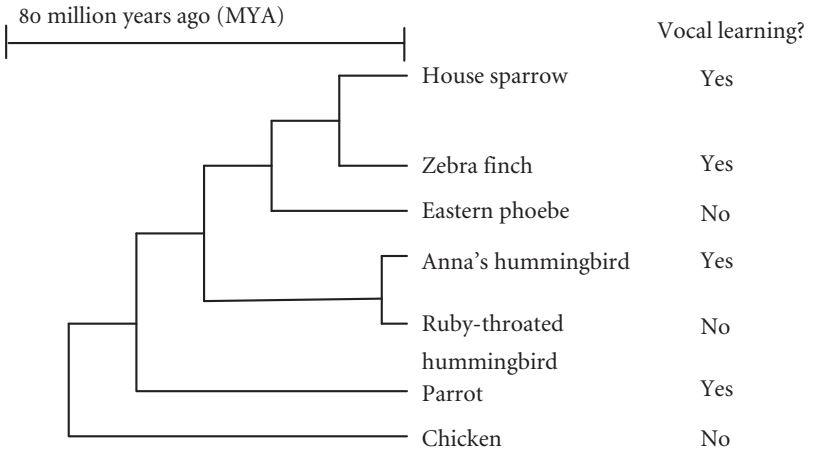


FIGURE 20.3 (Redrawn after Zhang, Webb, and Podlaha 2002; Jarvis 2004.) The pattern of vocal learning–non-learning traits in songbirds is considerably denser with respect to species than in the case of primates. For example, two very closely related species of hummingbirds (*Trochiliformes*), one Anna's hummingbird, the other, the Ruby-throated hummingbird, differ: Anna's hummingbird does vocal learning while the Ruby-throated hummingbird does not. On the other hand, a lineage that has a more distant most recent common ancestor to the hummingbirds, the parrots (*Psittciformes*) do exhibit vocal learning. This kind of pattern is much more suitable for comparative evolutionary analysis: we can at least raise the question as to whether vocal learning is an independent autapomorphy or not

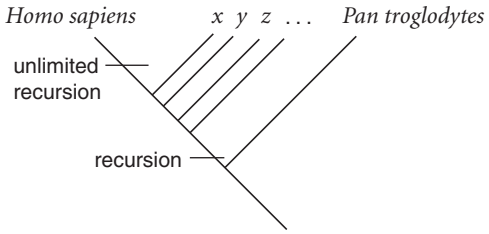


FIGURE 20.4 Notional (and fictional) cladogram with the trait of recursion marked as appearing at some point after the split with chimpanzees (*Pan troglodytes*). Here *x*,*y*,*z*,... denote a range of intermediate fossil species, which might or might not have had the trait of recursion or not (*Orrorin tugenensis*, *Australopithecus ramidus*, *Ardipithecus ramidus*, *Australopithecus anamensis*, *Australopithecus afarensis*, *Homo antiquus*, *Australopithecus bahrelghazali*, *Kenyanthropus platyops*, *Australopithecus africanus*, *Australopithecus garhi*, *Paraustralopithecus aethiopicus*, *Australopithecus aethiopicus*, *Paranthropus robustus*, *Australopithecus robustus*, *Australopithecus walkeri*, *Zinjanthropus boisei*, *Australopithecus boisei*, *Paranthropus crassidens*, *Australopithecus crassidens*, *Homo antiquus praegens*, *Australopithecus praegens*, *Homo habilis*, *Homo louisleakeyi*, *Pithecanthropus rudolfensis*, *Homo rudolfensis*, *Homo microcranous*, *Homo ergaster*, *Pithecanthropus erectus*, *Homo erectus*, *Homo antecessor*, *Homo heidelbergensis*, *Homo rhodesiensis*, *Homo helmei*, *Homo neanderthalensis*, *Homo sapiens*, and others to be described.)

non-learners in the avian lineage (many thousands of species are not shown in the diagram). This is the type of trait dispersion that is more closely tailored for comparative evolutionary study.

Finally, in Figure 20.4 we sketch a purely notional (and fictional) cladogram illustrating what one might hope to find in the case of the human-language FLN trait (here noted as "recursion"). The split between recursion and unlimited recursion is marked here as a representative example of what one would ideally like to find in the case of comparative evolutionary analysis where a trait appears on a lineage, as with vocal learning in birds; it is completely fictional.

Unfortunately, Figure 20.4 is more than speculative: it is almost certainly incorrect. As far as we know, the FLN is unique to the human lineage. If this is so, then what one can say about a unique autapomorphy is anything at all, or nothing at all. And saying nothing is much more compact. Note that this does *not* mean that one cannot run an adaptationist argument in such situations; just that it is more difficult. Of course, this has never stopped anyone in the evolution-and-language business, possibly including this author, since one can spin any kind of story whatsoever, and compelling story-telling is something our species does best.

Turning to a second major conceptual challenge of evolutionary theory's explanatory "problematic," there is the question of how to infer the operation of selective forces in the past, given that we have only data about the here-and-now. Once again we must face general issues as well as those particular to human language that make this inference more difficult than has sometimes been realized.

The general difficulties with evolutionary inference center on the distinctive character of evolution by natural selection: its time course and speed; its strength; and the possibility of interference from causes other than selection. Though demonstrably effective, as Darwin himself noted, natural selection operates over very long time scales. Typically, the selective advantage of even a highly beneficial gene variant is slight compared to the original, on the order of a tenth of one percent—that is, just one additional surviving gene variant out of one thousand. Further, the force of natural selection is easily overwhelmed by most other forces that can also alter gene frequencies—for example, demographic effects like migration. As Dobzhansky and other founders of the evolutionary Modern Synthesis of the 1930s realized, as few as one or two migrants per generation between populations can block two populations from drifting apart (one of the reasons for invoking 'reproductive isolation' as the definition of a species, despite its numerable problematic aspects). The upshot of these constraints is that the natural selection simply does not operate on a human time scale: we cannot see it happening on the wing, as it were, but are reduced to taking static snapshots of an ultimately long-term dynamical process, as Lewontin (2002) notes. Equation (1) below illustrates this in the simplest possible case of one gene having two variants, or alleles, with frequencies p and $(1 - p)$:

$$\Delta p = \frac{p(1-p)}{2\bar{w}} \frac{d\bar{w}}{dp} \tag{1}$$

From this equation we can immediately see that the amount of evolutionary change—the change in frequency of gene variant delta p—is directly proportional to the product of two terms: first, the variance of the gene variants in the standing population, $p(1 - p)$, at a particular time; second, the derivative (slope) of the natural logarithm of mean fitness with respect to the frequency p. Viewed this way, the equation has a natural geometric interpretation of evolution as "hill climbing" via gradient ascent through an adaptive landscape while always increasing mean fitness, in Sewall Wright's famous picture (1932), a precise form of some of the popular recent verbal expositions of this notion. The first term corresponds to the step size taken each generation, while the second term is the slope and direction of the ascent—which way the climb is

headed. (It is quite crucial to emphasize that equation (1) and its topographic interpretation holds only in the simplest possible case, with just two variants of a single gene, no interacting genes, and no interactions of fitness with frequency; as soon as such more realistic complications are introduced, the smooth rise to a single adaptive peak does not follow.)

Supposing now that a new gene variant starts out at some low frequency, say with p nearly 0, then evolutionary change will initially be very, very tiny; then it will gather steam, becoming greatest (steepest slope) when p and $(1 - p)$ are both at their half-way points, 0.5; after this, evolutionary change again levels off, asymptotically approaching zero as p fixes at frequency 1.0 and the variation in the population between the two gene types goes to 0.[3] Variance is thus the jet fuel that evolution burns—no variation, no evolution. The overall shape of the curve is sigmoid (a logistic equation), with an inflection point exactly at the middle. So unless one happens to catch change occurring at roughly the 50–50 midpoint, it will in general be difficult to see evolution in action. Further complications to this model may be introduced by adding the stochastic effects of finite population size, as is familiar, but we put to one side this important topic here.

An equation like the one above gives us a way of computing the evolutionary dynamics of what *will* happen going *forwards* in time, from some known starting point. However, this is not the inferential situation in which we find ourselves. Instead, we are only given some array of data—on frequencies of traits, genes, what have you—in the here and now. Our goal is to determine what the forces were in the past, including selection, migration, and so forth, that, starting from some unknown initial state at some point in the past, conspired to yield a trajectory with this observed end-state data. But as Lewontin (2002) observes, this is very nearly an ill-posed or under-determined problem with three unknowns and essentially just one equation: we cannot know the initial state; we do not know the forces that have operated over some (generally unknown) period of time—that is what we aim to solve. We know only the end-state. There is a standard escape to this problem in evolutionary inference, as Lewontin notes:

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. There is one solution to this dilemma. If the evolutionary process has gone on for a sufficiently long

[3] This over-simplified continuous form equation assumes an arbitrarily large population, so the frequency of p never actually reaches zero. In a more realistic form, with finite demographics, at some point p actually will go to zero.

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. What this means . . . [is that] all the effects of the initial state have disappeared. 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. (Lewontin 2002: 5)

20.1.2 *Case study: The problem with FOXP2 and the evolutionary explanation of language*

The recent and much publicized research on the *FOXP2* transcription factor gene serves as an excellent case study of the difficulties of evolutionary analysis, revealing the problems with backwards inference from observed present conditions, sensitivity to equilibrium assumptions, and conclusions drawn given a relatively small number of differences amongst a thinly populated species space. Apparently there are just two amino acid differences in this gene between *Homo sapiens* and Neanderthals.

Taken together with the apparent connection between defects in this gene and language disorders, the evolutionary analysis of *FoxP2* in other species and humans has sometimes been implicated as part of the push to language. We shall see that argument here is tendentious, both statistically and biologically, because when there are so few differences at such a far remove from a phenotype, and so much intervening time, it may simply be impossible to tell whether the gene in question was "selected for" or not, or when this selection happened, if at all. There are so many modeling parameters it resembles weather forecasting: assumptions have to be made regarding population size and changes, generation times, and selection coefficients, among others. This is well known in the evolutionary literature, but the consequences of changing the assumptions—a sensitivity analysis—are sometimes not explored. That is what we shall do here. The results point to a much weaker connection between *FoxP2* and language. While the evolutionary analysis remains cloudy, it may be that if we look at this gene in a different light, as part of the genomic machinery connected to externalization and serial motor coordination, then we can revive its analysis as a probe into language and evolution, a matter we take up in Act III, below.

To begin, let us quickly sketch the basics of the *FOXP2* story, its connection to language disorders, and its comparative evolutionary genetic analysis, omitting many details that may be found in several excellent summaries; we follow the presentation in Marcus and Fisher (2003). They describe the connection

between *FOXP2* and language this way (original references in the text have been omitted for clarity):

> The first direct evidence of a specific gene that influences speech and language acquisition has come not from complex traits, but from an unusual autosomal dominant form of communication disorder that is caused by mutation of the forkhead box P2 (*FOXP2*) gene . . . The consequences of *FOXP2* disruption differ from typical SLI [Specific Language Impairment, RCB] in that they include prominent difficulties in learning and producing sequences of movements that involve the mouth and lower part of the face. Affected individuals have problems with speech articulation (developmental verbal dyspraxia or DVD), which are accompanied by wide-ranging deficits in many aspects of language and grammar. Crucially, although general intelligence varies among individuals who carry the same *FOXP2* mutation, speech and language deficits are always evident, even for children with normal non-verbal intelligence. Moreover, the associated problems with processing language and grammar are not exclusively tied to speech—they are evident in the written domain and occur for comprehension as well as expression. . . . The link between *FOXP2* and disordered language was initially identified through genetic studies of a large three-generational family (known as KE), in which affected members carry a heterozygous missense mutation that alters the DNA-binding domain of the FOXP2 protein. The KE substitution markedly affects the function of the encoded protein (J. Nicôd, S.C. Vernes, F.M. Elahi, A.M. Coupe, L.E. Bird and S.E.F., unpublished observations). *FOXP2* mutations are not a predominant cause of language impairment; however, a second heterozygous point mutation in *FOXP2* was recently identified that co-segregates with speech and language deficits in another family. This nonsense mutation severely truncates the protein, deleting essential functional motifs, including protein–protein interaction domains, the DNA-binding domain and suspected nuclear localization signals. Independent chromosomal aberrations (including translocations and deletions) that disrupt *FOXP2* are associated with speech and language deficits. (Marcus and Fisher 2003: 5–6)

Turning now to evolutionary analysis, the crucial data has been provided by a phylogeny based on a comparative study of certain *FoxP2* regions from individuals of five different primate species along with an outgroup comparison to the mouse, as illustrated in Figure 20.5, redrawn after Enard et al. (2002). The crucial evolutionary distinction to focus on is the difference between what are called synonymous and nonsynonymous amino acid substitutions in these gene regions. A synonymous substitution is a change in one of the triplet DNA codon bases (Adenine, Cytosine, Thymine, Guanine) that does not alter the corresponding amino acid that is coded for—this is because the DNA triplet code is redundant, usually with several distinct triplets or "codons," especially those varying only in the third position, "spelling out" or specifying the same amino acid. Because a nonsynonymous change does not alter the corresponding amino acid comprising the protein coded for by

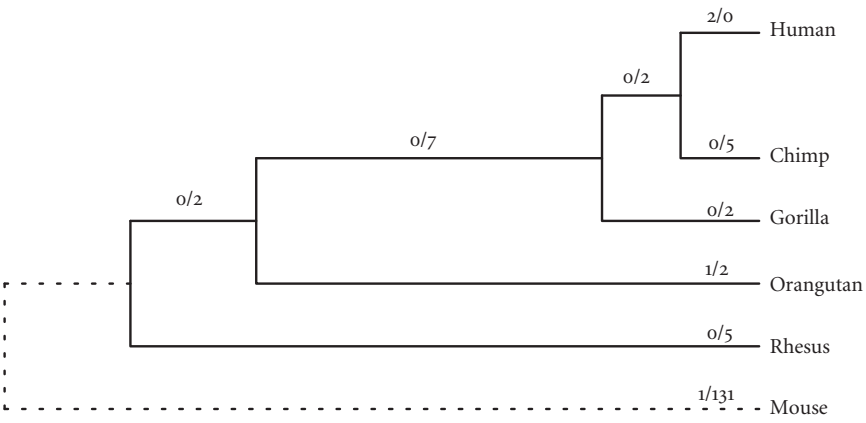


FIGURE 20.5 (Redrawn after Enard et al. 2002.) The phylogenetic relationship of FoxP2 exonic regions for five primate species (human, chimpanzee, gorilla, orangutan, and rhesus macaque) along with a comparison outgroup of *Mus musculus*(mouse). The upper/lower numbers denote the number of nonsynonymous (amino-acid changing)/synonymous (non-amino-acid changing or silent) substitutions along that branch of the lineage, respectively. For example, there are two nonsynonymous substitutions on the line leading from the least common ancestor of *Homo sapiens* and *Pan troglodytes* to *Homo*, at positions 303 and 325, and zero synonymous substitutions. The two changes were Threonine to Asparginine at amino acid position 303 in exon 7 and Asparginine to Serine at position 325. There are one nonsynonymous substitution and 131 synonymous substitutions on the line leading from the common ancestor of the primate clade and *Mus musculus* to *Mus.* See the main text for the results of computing a McDonald—Kreitman statistical test for detecting selection given these data, which turns out to be statistically indistinguishable from chance.

— ital

the *FoxP2* gene, it is assumed to be functionally silent and so not subject to direct natural selection. (Like many assumptions in biology, this one may actually not hold, but we put this issue to one side here.) For example, the amino acid Threonine (Thr) is spelled out by the DNA codon A–C–{T,C,A,G}. Therefore, any possible change to the third position to the DNA will not alter the corresponding amino acid. For example, a single nucleotide point mutation in the third codon position, from ACT to ACG will not result in any change in the corresponding protein coded for—it will remain Threonine. The amino acid Serine is redundant in the third position as well: it is spelled out by T–{T, C, A, G}. Serine has a further redundancy: it is also spelled out by the triplets AGT and AGC. To repeat, such changes are therefore assumed not to matter for selection, because they do not alter protein function.

In contrast, the amino acid Asparagine (Asn) is spelled out by the codons A–A–{T, C}, so any change in the third position from T or C to A or G will be *nonsynonymous* because it will result in a DNA codon that spells out a *different* amino acid from the original, namely, Lysine. Since such a change, a point mutation results in a new amino acid that may have functional consequences, it is assumed to be subject to selection. Note that both synonymous and nonsynonymous changes arise stochastically as the result of random mutational processes. By comparing synonymous to nonsynonymous changes, evolutionary biologists have developed a range of statistics to see whether natural selection has been operating, subject to certain assumptions.

The Enard et al. (2002) evolutionary study sequenced the genomic regions of exons 4–7 of *FoxP2* in five extant primate species (see Figure 20.5), along with mouse (*Mus musculus*), in order to carry out a comparative evolutionary analysis of exactly this sort. Enard et al. found that there were just two key nonsynonymous amino acid differences between non-human primates, mouse, and human: humans, but not the other species, have the DNA codon AAC specifying the amino acid Asn at position 303 in exon 7 of the gene, and DNA codon AGC specifying Serine (Ser) at amino acid position 325, just a bit further along down the gene. The corresponding DNA codons in chimpanzee *Pan troglodytes* and *Mus* are both ACC, specifying the amino acid Threonine (Thr), and AAT, specifying the amino acid Asparginine (recall that Asp is redundantly specified by several different triplet codons). This evidence suggests that the ancestral codon state was AGT, and a single nucleotide mutation in the second position from A(denine) to G(uanine) changed this to AGC, and so altered the amino acid coded for at position 303, a nonsynonymous change; similarly, that ACC was ancestral and mutated to AGC via a single nucleotide change from C(ytoseine) to G(uanine), changing the amino acid at position 325. (Neither of these changes are those involved in the KE-family disruption, which are due to other defects in *FOXP2*.)

We should also note that this analysis assumes by convention that just a single nucleotide change has occurred here, say from A to G—an assumption of parsimony. Of course it could also have been possible that the A mutated to G, then back to A, then to G, and so on, along the line leading from the common ancestor of mouse and human to human. One simply cannot know for certain; it is merely simplest to assume that there was but one such change. Lewontin (1989) notes this is true simply because, first, "any nucleotide position has only four possible states, so that although two sequences may be *observed* to be identical at some position, they may be separated by numerous

evolutionary changes but have converged to the same state," and, second, "each state is, at least mutationally, accessible from each other state so sequences of evolutionary states are not well ordered" (1989: 15). Lewontin then points out that how to correctly estimate the number of nucleotide substitutions along an evolutionary lineage can radically differ, depending upon whether amino acid substitutions (nonsynonymous changes) are completely unconstrained, strongly constrained, or somewhere between these two extremes. In the case of *FoxP2*, there is a (perhaps tacit) assumption that amino acid substitutions have a strong functional impact, but it is not clear that this was taken into account (note that except in the case of mouse, we have very little evidence of a large number of synonymous substitutions as compared to nonsynonymous substitutions for any of the primate species, the hallmark of strong functional constraints on amino acid substitutions, as Lewontin notes.)

Putting such questions to one side for the moment, let us examine the phylogenetic diagram in Figure 20.5 more carefully. Time runs left to right. The horizontal line after each branch point between two species is labeled with two numbers, one above the other. The top one denotes the number of *nonsynonymous* amino acid changes in exons 4–7 after the branch point, the bottom denotes the number of *synonymous* amino acid changes in exons 4–7. So for example, in the Enard et al. sequencing data there is one nonsynonymous amino acid change in mouse and 131 synonymous substitutions after the common ancestor between all the other primates and mouse. Similarly, since the split between chimpanzees and human, seen there have been two nonsynonymous amino acid changes in humans, and there are apparently zero synonymous substitutions, for a set of individuals drawn from a wide variety of geographic regions.

From this data, Enard et al. drew a number of striking evolutionary inferences. Perhaps most importantly, they computed a number of standard statistical tests to detect natural selection, suggesting that there was a "selective sweep" in the case of *FOXP2*—that is, strong positive selection for specific *FOXP2* changes along the evolutionary line that led to humans, as well as an estimate of the time when that sweep occurred, perhaps 50,000–100,000 years BCE.

But are these strong conclusions justified? There seem to be at least three main difficulties:

1. The tests that were used to detect selection at the two crucial amino acid differences between us and primates adopted the standard assumption of 'stochastic equilibrium' as described above. A re-analysis using the only statistical test known that is *not* sensitive to such an assumption (McDonald

and Kreitman 1991; Kreitman 2000), details given below, reveals no statistically significant positive selection.
2. The simulation study Enard et al. used to estimate the time when this putative selection acted also made certain biological assumptions about the strength of natural selection, generation times, and particular computational assumptions regarding numerical accuracy. When we alter these, for example, changing selection to a more naturally occurring value, then the sweep disappears.
3. Even supposing that *FOXP2* was "selected for," it remains unclear exactly *what* function it might have been selected for: initially it might not have been selected for its role in serial motor control, always an issue for a transcription factor gene that regulates other genes. At this point we simply cannot say.

Let us take up these points one by one. First, consider the assumption of stochastic equilibrium. From the data in Figure 20.5 one can compute a two-way contingency table, the McDonald–Kreitman statistical test for detecting selection. The idea is to compare the number of synonymous and nonsynonymous amino acid substitutions *within* species and also *between* species, for example, both *within* humans and *between* humans and, in this case, mouse. By doing a comparison both within and across species groups and looking at the ratios, the intuition is that any stochastic ripple that could affect one column would also proportionately affect the other (roughly because all historical events that could have jiggled both groups' numbers would have affected both in equal proportion; for further explanation, see McDonald and Kreitman 1991). This is the only test for selection we currently have that does not rely on the assumption of stochastic equilibrium (Kreitman 2000). There is a price paid for this robustness: the test is known to be very conservative, that is, it errs on the side of rejecting true positives. Thus, if one has a McDonald—Kreitman test that says that selection has occurred at some statistically significant level (say $p = 0.05$ or 0.01) then one can be fairly confident that selection has indeed taken place. Of course, the converse is not true; but as we have noted, all the stronger tests are subject to the slings and arrows of the equilibrium assumption.

What about the case at hand? There are $1 + 2 = 3$ between-species nonsynonymous amino acid differences between *Homo* and *Mus*, and $131 - 3 = 128$ between-species synonymous differences. Enard et al. found 47 within-species *Homo* synonymous amino acid differences across the different individuals they sampled, and 0 nonsynonymous differences. This yields the two-way contingency table in Figure 20.6, for which we can use Fisher's exact test to

	Human	Mouse
Synonymous amino acid substitutions	47	128
Nonsynonymous amino acid substitutions	0	3

FIGURE 20.6 A McDonald–Kreitman (two-way contingency table) test on the between- and within-group synonymous/nonsynonymous *FoxP2* exonic data from Enard et al. (2002). A calculation of the probability of such a pattern appearing solely by chance using Fisher's exact test finds that the probability that this array of counts could be due to chance alone to be in effect 1.0, i.e., near certainty. Thus it is highly statistically insignificant (p value effect in 1.0, i.e., at chance level.

find the precise probability that this distribution of numbers could be due to chance alone, which happens to be very nearly 1.0. The reason for this negative result is that there are so few nonsynonymous differences between *Homo* and mouse (or chimpanzee), and no nonsynonymous variation at all within *Homo*.

Second, the simulation study Enard et al. used to estimate the time when this putative selection acted also made certain biological assumptions about the strength of natural selection, generation times, and particular computational assumptions regarding numerical accuracy. In particular, Enard et al. found that a selection coefficient of 0.01 (1%) yielded a statistically significant selective sweep under their model, but importantly, note that when this value is reduced to a value that has more often been found in field studies of selection, $\frac{1}{2}$ to $\frac{1}{10}$ of a percent, then the sweep disappears: in fact, these more biologically realistic values about the strength of selection (as Enard et al. themselves note) do not lead to the same statistically reliable results.

Third, even supposing that *FOXP2* was selected for, it remains unclear exactly *what* function it might have been selected for: initially it might not have been selected for its role in serial motor control, always an issue for a transcription factor gene that regulates other genes. We simply do not know. While *FOXP2* clearly operates in modern human neural development, and its disruption affects motor learning and language in human and other species, as confirmed by an increasing number of studies (e.g. Teramitsu et al. 2004; Haesler et al. 2007), it is also presumptively involved in the formation of the digestive gut epithelial lining. Crucially, the key amino-acid change proposed

as being under strong positive selection, at position 325, also seems common to all *Carnivora* (see Figure 20.7). If so, then the human *FOXP2* changes might well have been due to dietary modifications related to the move from trees to savannah, with the related neural effects a concomitant effect, in this sense a kind of hitchhiking. The causal sequence remains unclear.[4]

Summarizing, what should we conclude about the evolution of human language and the observed evolutionary changes in *FOXP2*? If anything, this re-analysis serves as a cautionary example of the challenges of evolutionary inference when there are so few differences between closely related species. Some confirming evidence for this conclusion comes from the first comparison of the human genome and the draft chimp genome, as reported by the Chimpanzee Sequencing Consortium in *Nature*, 2005. This study compared 13,454 "matching" (orthologous) human–chimp genes, and found evidence for accelerated evolution in the human lineage for only 585 genes, all but a handful related to the expected categories of immunity, olfaction (humans lost their sense of smell), and reproduction (e.g. spermatogenesis).

The Consortium also carried out comparisons of these genes and others in mouse and rat with respect to synonymous vs. nonsynonymous substitution reported in their Supplementary Data (details were omitted in the main text). In contrast to other studies, *FOXP2* did not stand out: the ratio of synonymous/nonsynonymous substitutions for *FOXP2* in human was 0.81, a ratio one expects to see when there is no selection going on.[5] Again, this is not a surprising result; it is simply what happens when we look for differences by examining species that are too close, or, as the Consortium put it in the case of *FOXP2*: "given the small number of changes involved, additional data will be required" (2005: 79). The *FOXP2* changes could have been due to chance alone. What can be done about this situation? Act II suggests the obvious course: look at species that are farther away. Since there is increasing evidence that the

[4] A final, more minor point is that Enard et al. provide some evidence that they suggest points to a *functional* reason why the amino acid under possible positive selection might have changed in the manner it did: a computer simulation indicating that this change prompts an alteration in the gene product's binding site. Repeating their analysis with three other tools confirms this. However, there is one problem: in other species that have also been argued to have accelerated *FoxP2* evolution (bats, birds), this particular amino acid is *not* changed. Rather, the changes occur at other positions in the DNA sequence of *FoxP2*, and when one carries out the same computer simulations, there appears to be no comparable functional change in these other species. I conclude that this functional evidence is weak at best.

[5] For comparison, the Consortium's ratios for mouse and rat were both 0.15, a number indicative of so-called negative or purifying selection. The chimpanzee ratio could not be tested, as the Consortium actually found zero substitutions of either sort.

Order	Species	Site 303	Site 325
Galliformes	Chicken	Thr	Asn
Tubulidentata	Aardvark	Thr	Asn
Artiodactyl	Pig	Thr	Asn
	Cow	Thr	Asn
Cetacea	Whale	Thr	Asn
Perissodactyla	Zebra	Thr	Asn
	Tapir	Thr	Asn
Carnivora	Cat	Thr	**Ser**
	Dog	Thr	**Ser**
	Wolf	Thr	**Ser**
	Wolverine	Thr	**Ser**
	Bear	Thr	**Ser**
	Fox	Thr	**Ser**
	Sea lion	Thr	**Ser**
Chiroptera	Bat	Thr	Asn
Rodentia	Mouse	Thr	Asn
Lagomorphs	Rabbit	Thr	Asn
Insectivora	Mole	Thr	Asn
Primates	Lemur	Thr	Asn
	Tamarin	Thr	Asn
	Rhesus	Thr	Asn
	Gorilla	Thr	Asn
	Chimp	Thr	Asn
	Bonobo	Thr	Asn

FIGURE 20.7 (Excerpted and redrawn from Zhang et al. 2002, Figure 3.) The bounded box highlights species where position 325 of the *FoxP2* transcription factor gene codes for the amino acid Serine (Ser) rather than Asparginine (Asn), the same putative change under selective pressure in *Homo* according to Enard et al. (2002). Note that this encompasses the entire order *Carnivora*, as reported by Zhang 2002.

transcription factor protein FoxP2 assists in neural development, including the development of serial motor coordination, specifically in vocal externalization and imitation in birds, then perhaps we can rescue the FoxP2 story by resurrecting it in another guise—not as the hallmark of recursive syntax, but as part of how serial motor externalization (and possibly vocal imitation) are linked. If so, then *FOXP2* would not speak so much to the origin of the core of human language, recursive syntax, but to how language is externalized. Act II examines this possibility in more detail, by considering the metrical structure of language.

20.2 Act II: Merge from the Bottom Up—The Return of the FOX?

According to the Minimalist Program, we can envision the entire computational system associated with language as having two main interfaces with the other cognitive/computational systems external to language proper: the first the conceptual–intentional interface, roughly, the interface between syntax and the systems of thought, belief, reasoning, and the like; the second the sensori–motor and articulatory–perceptual interfaces comprising the connection between syntax and its external form (either its perception, via parsing, or its production as, for example, the articulatory gestures of spoken language or sign language). The central operation of this central computational language system, or "CS" as it is dubbed by Reinhart (2006), is the single operation Merge.

Act I reviewed some of the existing comparative evolutionary evidence and arguments available that might shed light on the natural selection for Merge, finding these lacking, primarily due to the autapomorphic (species-specific) nature of Merge and issues with primates being "too close" to humans and too sparsely populated in species space. Act II attempts to remedy these problems by taking a highly speculative and quite radical position: instead of studying species that are quite close to humans, like chimpanzees or other primates, perhaps one can find an analog in another domain, in a species at some remove from humans so that the problem of accidental convergence does not arise. Act II argues that such an analog might be found in the process by which the metrical structure of sentences—their rhythmic character—is formed, and that this process may well be shared with songbirds. If this approach is on the right track, then one could use comparative evolutionary analysis in birds as a window into similar processes in humans. In particular, Act II adopts the Halle–Idsardi model of metrical stress assignment, as refined in Fabb and Halle (2006), as its theory of metrical stress assignment. It shows how this model, essentially a form of counting, applies to human metrical

patterns, and then extends that analysis to songbirds using a novel method for extracting syllables posited by Coen (2006). The end result is a skeletal form of Merge: it operates without formal lexical features of the sort found in syntax proper, combining only the asterisks that Fabb and Halle associate with basic syllables. In this sense, it is pure syntax—literally a skeleton—with the combination-and-selection character of Merge (or formerly X-bar theory), but without any other features at all of the sort usually associated with lexical entries. But this is just what one would like for comparative purposes, since presumably songbirds lack precisely these formal syntactic features as well. Birds have songs, but no semantics or lexical items in the conventional sense. Primates face the opposite problem: they seem to have a conceptual–intentional system, perhaps even lexical items ("words") according to much evidence accumulated over the past several decades, but seemingly lack the Merge operation itself (otherwise, they would indeed have the capabilities of human language). Only humans have lexical items and Merge, yielding a fully recursive syntax.

To proceed, we first consider how the Halle system for the assignment of metrical stress or rhythmic structure to syllables works. We then apply it to the songbird case. Finally, we show how this really does amount to a reduced form of Merge.

To begin, we sketch the Halle system, drawing directly from the exposition in Fabb and Halle (2006). The process works by the repeated counting and grouping of syllables, denoted simply as asterisks, according to an ordered set of rules, forming as output a metrical grid whose periodic structure reflects whether a syllable will be stressed or unstressed. An example taken from Fabb and Halle (2006) will serve to illustrate. Suppose we have the English line, *Tell me not in mournful numbers* (from the poem by Longfellow). Fabb and Halle present an ordered list of seven rules (their 12a–g) that accomplishes this. We can mark out the syllables as follows, assuming the following associated metrical structure where a slash marks a syllable likely to have strong stress, while an "x" marks a syllable likely to have weak stress or no stress, where we have inserted dashes for readability:

/ x / x / x / x
Tell me not in mourn-ful num-bers

Fabb and Halle's rules associate an initial set of asterisks ("gridline 0") with the syllables in this line, and then their ordered rules operate to produce a vertical grid that matches this rhythm. Initially the asterisk marking, one per syllable, is given by their first rule, (12a), "Project each syllable as an asterisk on gridline 0," as follows:

Tell me not in mourn-ful num-bers
* * * * * * * *

Next, Fabb and Halle's rule (12b) groups these asterisks into pairs by inserting left parentheses, left to right, obtaining what Fabb and Halle call 'the gridline 0 feet' (we have again inserted dashes for readability):

Tell me not in mourn-ful num-bers
(* * (* * (* * (* * gridline 0

We now apply rule (12c): the leftmost asterisk in the next group is projected to the next gridline (gridline 1):

Tell me not in mourn-ful num-bers
(* * (* * (* * (* * gridline 0
* * * * gridline 1

We now apply their rule (12d) on gridline 1, moving now from right to left every two asterisks and inserting a right parenthesis to obtain a new grouping:

Tell me not in mourn- ful num-bers
(* * (* * (* * (* * gridline 0
* *) * *) gridline 1

We next apply their rule (12e): the rightmost asterisk on gridline 1 is projected to form the next gridline, gridline 2:

Tell me not in mourn-ful num-bers
(* * (* * (* * (* * gridline 0
* *) * *) gridline 1
* * gridline 2

We now re-apply rule (12d) (rule (12f)) to gridline 3:

Tell me not in mourn-ful num-bers
(* * (* * (* * (* * gridline 0
* *) * *) gridline 1
* * *) gridline 2

Finally, we apply the rightmost projection rule one more time, their (12g), obtaining gridline 3 and the final result (since there is only one asterisk remaining and no more rules can apply):

```
Tell me  not  in mourn-ful num-bers
(*  *  (*  *   (*   *   (*   *  gridline 0
 *      *)      *       *)      gridline 1
         *                *)     gridline 2
                          *      gridline 3
```

If we now match up the grid positions with the most asterisks, we see that they correspond to the observed rhythmic pattern of the line as desired:

```
/    x   /   x    /    x   /    x
Tell me  not  in mourn-ful num-bers
(*  *  (*  *   (*   *   (*   *  gridline 0
 *      *)      *       *)      gridline 1
         *                *)     gridline 2
                          *      gridline 3
/    x   /   x    /    x   /    x
```

Of course, this is just one line of words from a single language, while there are many metrical patterns and many languages. According to Fabb and Halle's analysis, all of the various valid metrical patterns of human languages may be defined by parameterizing their procedure above as a list of binary choices for each of the gridlines. We list their parameters as follows (0/1 numbers denote our binary encoding of the parameter values):

1. Insert either left or right parentheses (0/1).
2. Insert parentheses at every binary/ternary asterisk (0/1).
3. Start inserting parentheses at either the left/right edge of the grid (0/1).
4. Project a "head" (an asterisk at the next gridline) by selecting either the leftmost or the rightmost asterisk of a group delimited by parentheses and placing an asterisk above this one on a new line above the existing line (0/1).
5. Start the process at either the first or the second syllable (0/1).

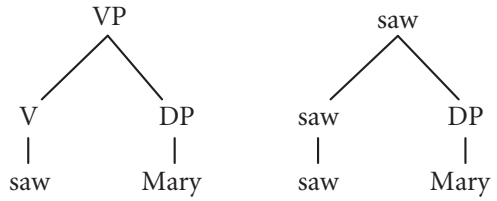
In the Longfellow example, we can see for instance that the parameterization applied to the first line of the grid was to (a) insert left parentheses; (b) insert binary; (c) start at the left; (d) project left; (e) start at the first syllable. Thus as a bit-vector this parameterization would be simply [0 0 0 0 0]. Note that at the second grid line, we reversed direction, and inserted right parentheses starting from the right, promoting the rightmost asterisk to the next gridline, for instance, parameterization [1 0 1 1 0]. Crucially, Fabb and Halle note that certain parameterizations will lead to invalid metrical structures, that is, those that do not seemingly correspond to the particular rhythmic structure of

978–0–19–955327–3 20-Sciullo-c20 Di-Sciullo (Typeset by SPi, Chennai) 485 of 491 May 11, 2010 13:25
OUP UNCORRECTED PROOF – FIRST PROOF, 11/5/2010, SPi

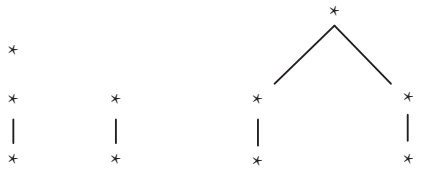
this example. For instance, as they note, if instead the system moved from right to left at gridline 0, the resulting grouping and projection of a head would result in an improper pattern, where the first syllables of *mournful* and *numbers* would not be placed on gridline 1, to be eventually stressed, which is incorrect:

```
/    x  /  x    /    x   /   x
Tell me not in mourn-ful num-bers
*    *) *  *)   *    *)  *   *)  gridline 0
     *     *        *       *     gridline 1
```

Though the Fabb and Halle system is presented in a form that uses the vocabulary of selection, projection, and heads (as in X-bar theory), we may recall that in the Minimalist Program, this vocabulary has been replaced by the (largely minimally necessary) properties of Merge. We can illustrate this by comparing an example from the older X-bar system in syntax, and then a corresponding example from the Fabb–Halle gridline system. In the systems antedating Merge, we would say that a verb phrase *saw Mary* is comprised of the element *saw* (putting to one side tense, etc.), and the DP *Mary*. As is familiar from X-bar theory, the +V lexical item *saw* has been selected as the head of the phrase, and then projected to the next level, that of a constituent (the verb phrase), with all its features copied to this position, leading to the familiar hierarchical structure depicted below on the left. Indeed, some would go so far as to say that all the features of *saw* are copied to this position, leading to the second figure below on the right, where we place *saw* at the root node of the hierarchical structure formed by the combination of *saw* and the DP:



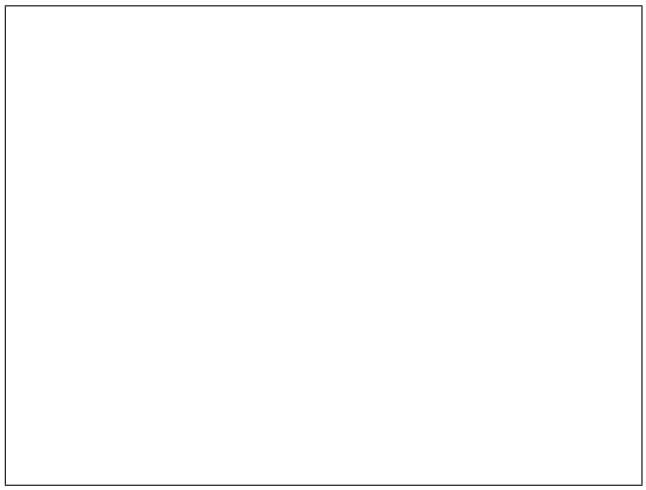
Of course, all this means is that "saw" and the DP have been grouped together, just as in the Fabb and Halle system, with "saw" selected as the head of the phrase and projected, as is familiar. Applying this to the asterisk-based system, we can write the Fabb and Halle notation as on the left, and its equivalent in X-bar form in the notation on the right.



It is easy to see that the Fabb–Halle system operates formally precisely like the familiar X-bar system, but with one crucial difference: there are no lexical features whatsoever, that is, features like +V associated with lexical entries, no agreement features, and the like. There are simply the marks as we have shown them in the diagram, which have no features. Nonetheless, the rest of the formal apparatus remains the same and operates geometrically as before: in both cases, one of the items is selected to be the head of the next line, and whatever features it has (in this case, the empty set, there being no features) are copied to that level.

A Merge-based account operates similarly. In its simplest form, Merge takes two objects, here just asterisks, and combines (i.e. groups) them into a single new object, selecting one as the head to be the label of the grouped object. But this again simply describes the basic operation of the Fabb–Halle selection-and-grid-projection procedure. Fabb and Halle's system differs only insofar as the label contains simply a single vacuous feature, namely, just an asterisk (and so probe–goal agreement applies vacuously). We might therefore regard it as the simplest (and degenerate) kind of system that exhibits the most basic property of grouping, that is, Merge. Indeed, Chomsky (2007) notes that Merge operating even more degenerately on just a single featureless item would yield counting (the number system), but without grouping. Thus the proposal advanced here is in this sense the simplest nontrivial extension of the Chomsky (2007) proposal, showing that metrical stress assignment, too, falls under the Merge model.

If this is on the right track, then we can ask the next question: this computational system, like syntax, must interface to the outside world via some externalization process. The final output from the Syntax–SM externalization process is assumed to have at least two properties: first, it must be linear (rather than hierarchical), that is, the syntactic objects passed from syntax are flattened so that the relation of precedence is imposed (a topic pursued in great depth in Kayne 1994); second, it is ordered left-to-right, in virtue of the necessary time course of sound and articulators operating in the real world. We unpack the externalization mapping into two steps that correspond to each of these properties: (1) impose precedence (what comes next to what, in the



AU: supplied as .pdf, .eps, or .tiff file

Au: Please provide the figure 20.8.

Externalization as a two-stage process

FIGURE 20.8

linear order projection); (2) take the precedence-ordered object and determine whether it is to be output left-to-right or right-to-left.

(See Figure 2.8.)

Minimally, this requires adding two new relations: (1) precedence, that is, whether one element comes before or after another (the minimal relation needed for any linear ordering); and (2) global order (whether the resulting sequence output right-to-left or lefttoright; note for example, that the grid-lines can be read in either order so there is indeed a choice). The output result is a series of high-to-low stress assignments. This suggests that much of the Fabb and Halle system might be forced by conditions at the external interface, leaving only binary Merge as its central operative principle, the best possible result.[6]

[6] We leave for future research this possible redundancy in the Fabb–Halle parameterization that might follow from this externalization procedure and its imposition of both precedence and left-to-right relations. Examining the Fabb–Halle parameters (1)–(5) it may be noted that many are left-to-right symmetrical: one can insert either left or right parentheses; one can proceed either from the left- or right-hand side; one starts inserting parentheses (counting) at either the right or left edge; etc. If, however, there is no "right" or "left" at some earlier internal stage of grouping that is simply Merge before externalization, then these parameters cannot be distinguished and may be collapsed together. This suggests that the system might be simplified by breaking it down into a two-stage process: stage 1, Merge of items with no lexical features; stage 2, externalization and thereby imposition of "left" and "right" parentheses (precedence) as well as externalization in left-to-right or right-to-left order, a condition imposed by the interface. If so, then the interface condition forces all of the parameterizations in (1)–(5) above, leaving just binary Merge. (We put to one side here the question about reducing Fabb and Halle's ternary counting, which might be derivative from binary Merge plus adjacency: on this account, the notion 'ternary' is a binary grouping, supplied by Merge, plus the addition at either the right- or left-hand edge, of a singleton. It remains to work out the details of this reduction.)

Since the Fabb–Halle system is based solely on counting syllable marks (asterisks) and their groupings, one might inquire as to whether the observable externalization of such a system of metrical patterns can be found in the vocalizations in other species. If so, perhaps we can use this as evolutionary insight into at least part of the language system. The answer here seems to be yes, at least for a preliminary set of cases. Consider birdsong, in particular a species that does vocal learning and where FoxP2 is also known to be involved, zebra finches, *Taeniopygia guttata.* The first issue that arises is what counts as a "syllable." The avian biologist's traditional division of a finch song into syllables would delimit then by silences—the start of a syllable is a period of silence, and similarly its end is marked by a period of silence. However, this does not correspond to what would be a linguist's analysis.

We can more properly approach the linguistic representation by applying a method developed by Coen (2006) to find what he calls "songemes." Coen processes a finch song by means of peak-power filtering, which produces a more nuanced division into regions of high and low intensities, rather than just regions of silence/no silence. This becomes a presumptive proxy for songemes: each looks like a hill—a rise followed by a drop in peak power. We can partition these as illustrated in Figure 20.9, from Coen (2006), where the blue vertical lines indicate the songeme boundaries, and the yellow curve highlights the peaks and valleys of the resulting metrical structure.

For example, we now consider in more detail the interval in Figure 20.9 from approximately 600 msec to 850 msec. In this case one can recover a pattern of peaks and valleys that amounts to roughly the following sequence of High–high–Low–low regularities, a metrical structure close, though not identical, to the Longfellow structure, H–L–H–L–H–L . . . We have analyzed several dozen finch songs available from the datasets in Coen's thesis, and they all fall under the Halle-type analysis, using parameters (1)–(5), usually trochaic or iambic in format: a Merge-type grouping system, followed by linear externalization, if the previous analysis is correct.[7]

From an evolutionary standpoint, this result, if correct, leads to the following connection. There has been recent accumulated evidence that interference with FoxP2 in these birds (via the knock-down of the *FoxP2* genes) disrupts syllable structure (on the production side, that is, either directly on motor externalization, or through some more complicated feedback cycle in terms of vocal learning and imitation); syllables are over-extended, doubled, or absent. Put more generally, syllabic metrical structure becomes defective

[7] Very preliminary analysis of humpback whale and other cetacean songs such as bluenose dolphins, for example, as in Suzuki, Buck, and Tyack (2006), yield similar results, as do subsonic elephant songs and ultrasonic mouse song.

FIGURE 20.9 A partitioning of a zebra finch song by means of a peak-power filtering. Note the hills and valleys. Vertical blue lines at maxima and minima divide the song into a more nuanced pattern of high and low intensities. (Courtesy of Michael Coen, from Coen 2006, figure 5.14.) Time is on the *x*-axis; frequency in kHZ on the *y*–axis; power is denoted by color intensity

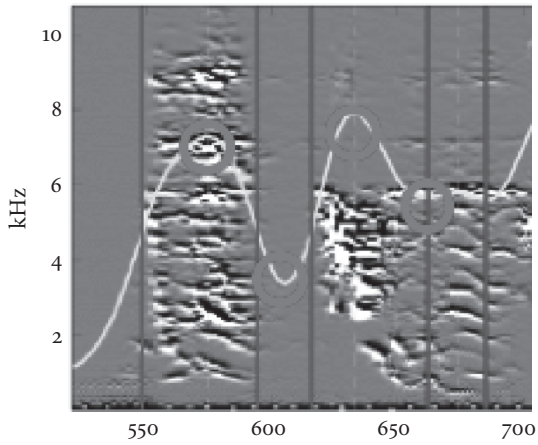


FIGURE 20.10 The high–low "metrical structure" of one section of the finch birdsong from Figure 20.9

(Haesler et al. 2007). While it is impossible to say in any detail exactly what aspect of the machinery has been disrupted, this result is consistent with the view that the human FOXP2 transcription factor, more generally the FoxP2 vertebrate transcription factor, is largely part of a secondary externalization process. However, there remains one (highly speculative) connection back to the internal system and Merge: if it is true that rhythmic structure is initially Merge-generated, without a lexicon, and then linearized at the S–M interface, then it might be that this computational ability shared by both birds and humans, and possibly the entire vertebrate lineage. Of course, it might also be true that this is an independent apomorphy, invented separately by both birds and humans to solve the interface problem for rhythmic structure. Only a much more detailed investigation of metrical structure in birds, as well as mammals, could begin to answer this; there are many open questions, among them whether the Merge-like complexity of the Halle system is required at all to describe or explain rhythmic structure and how this structure varies between vocal learners and non-learners. For example, if it is required only for vocal imitative learning, as suggested in Haesler et al. (2007), then it would surface only in the case of vocal learners. It also might well be true that this interface solution is entirely a third-factor constraint that has nothing whatsoever to do with Merge; it is simply a common solution to a similar functional problem, as in the similar aerodynamic shape of bird and bat wings. So while this "zero feature" Merge computational ability might be antecedently present in other, distantly related species, it raises both opportunities and puzzles: opportunities for evolutionary investigation in that it does

not encounter the problems with comparisons to other primates, but puzzles in that as Chomsky (2007) notes, unary feature-less Merge leads to arithmetic and counting. If so, why do we not observe the same arithmetical ability in birds as in humans? We leave all these as open questions, noting however that Chomsky (2007) asserts that unbounded Merge is associated with what he calls "undeletable edge features," rather than a totally feature-free lexicon as in the assumptions regarding rhythmic structure.

20.3 The Big Bang and a Venetian Fairy Tale: All You Need is Merge

This chapter has advanced three main points about the connection between evolution and language. First, it advocated caution: one can no more do armchair evolutionary theory than one can do armchair linguistics. Many folk intuitions about evolution and the ease of inferring that natural selection has operated in one or another scenario may simply be wrong, even for experts. In particular, echoing Lewontin (1998) but in the new domain of comparative genomics, inference in the sparse space of primate species is exceedingly difficult.

Second, it argued that where the FoxP2 transcription factor seems to be operating is in the domain of the S–M interface, but with a hint of Merge behind it, in particular, suggesting that the Fabb and Halle system for building metrical structure just *is* Merge, operating without lexical features and under the constraints of the external S–M interface system.

Third, and perhaps the most fanciful point of all, hence a Venetian fairy tale: it is suggested that if the story about metrical structure is correct, it allows us to rescue the FoxP2 story in a dramatic, though perhaps fanciful way to account for the Big Bang or apparently saltational appearance of human language. In line with what this chapter has indicated, Chomsky (2007) has sometimes advanced the view that what it took to get us language was in fact a small change, "at the minimum, some rewiring of the brain, presumably a small mutation or a by-product of some other change, provided Merge and undeletable E[edge] F[eatures] (unbounded Merge), yielding an infinite range of expressions." It may be possible to say more than this. Chimpanzees, and possibly other primates, got conceptual atoms. Birds got rhythm. But only people combined both, getting undeletable features, *and* Merge. And from this, came language: Merge is all you need.