4

Syntax Facit Saltum Redux: Biolinguistics and the Leap to Syntax

ROBERT C. BERWICK

I have been astonished how rarely an organ can be named, towards which no transitional grade is known to lead. The truth of this remark is indeed shown by that old canon in natural history of *Natura non facit saltum*. We meet with this admission in the writings of almost every experienced naturalist; or, as Milne Edwards has well expressed it, nature is prodigal in variety, but niggard in innovation. Why, on the theory of Creation, should this be so? . . . Why should not Nature have taken a leap from structure to structure? On the theory of natural selection, we can clearly understand why she should not; for natural selection can act only by taking advantage of slight successive variations; she can never take a leap, but must advance by the shortest and slowest steps. (Darwin 1859: 194)

4.1 Introduction: Language, Biology, and the Evolution–Language Gaps

For hundreds of years, long before Darwin's publication of *Origin of Species* (1859) and his *Descent of Man* (1871) that explicitly brought language into the fold of modern evolutionary thinking, the evolution of language has captured the imagination of biologists and linguists both. Among many evolutionary puzzles, one that stands out is the obvious discontinuity between the human species and all other organisms: language is evidently unique to the human lineage—being careful here to define language properly as distinct

This research was supported by NSF grant 9217041-ASC and ARPA under the HPCC program. Noam Chomsky, Samuel Epstein, Anna-Maria Di Sciullo, Charles Yang, and Morris Halle provided many valuable comments; all remaining errors are our own.

978–0–19–955327–3 04-Sciullo-c04 Di-Sciullo (Typeset by SPi, Chennai) 66 of 99 May 10, 2010 13:46
OUP UNCORRECTED PROOF – FIRST PROOF, 10/5/2010, SPi

from general communication, a matter addressed in Berwick and Chomsky (this volume) and considered further below.

Such gaps, or evolutionary novelties, have always posed a challenge to classical Darwinian analysis, since that theory is grounded fundamentally on the notion of gradualism—incrementally fine steps leading from a trait's precursor, with adaptive, functioning intermediates at every step along the evolutionary path, ultimately culminating in an "organ of extreme complexity and perfection," as with the vertebrate eye. Add one extra layer of light-sensitive membrane, so the argument goes, and an eye's photon-trapping improves by a fractional percent—a smooth incline with no jumps or surprises. Indeed Darwin himself devoted considerable effort in *Origin* to analyzing precisely this problem with this assumption (Chapter VI, "Organs of extreme perfection"), using the evolution of the eye as his make-or-break case study for his gradualist model, the very heart and soul of the theory itself, since, as he himself insists, "If it could be demonstrated that any complex organ existed, which could not possibly have been formed by numerous, successive, slight modifications, my theory would absolutely break down" (1859: 189):

> To suppose that the eye, with all its inimitable contrivances for adjusting the focus to different distances, for admitting different amounts of light, and for the correction of spherical and chromatic aberration, could have been formed by natural selection, seems, I freely confess, absurd in the highest possible degree. Yet reason tells me, that if numerous gradations from a perfect and complex eye to one very imperfect and simple, each grade being useful to its possessor, can be shown to exist; if further, the eye does vary ever so slightly, and the variations be inherited, which is certainly the case; and if any variation or modification in the organ be ever useful to an animal under changing conditions of life, then the difficulty of believing that a perfect and complex eye could be formed by natural selection, though insuperable by our imagination, can hardly be considered real. (1859: 186)

From this perspective, human language seems to be exactly the kind of unwanted biological surprise Darwin sought to avoid. Indeed, it apparently stands squarely as a counter-example to his entire theory of evolution via descent with modification, at least if we are to take Darwin at his word. Perhaps that is why ever since Darwin nearly all researchers seem to abhor even the slightest hint of a non-gradual, non-adaptationist account of human language evolution.

It is perhaps worth recollecting that this picture of evolution-as-minute-accumulation of small changes was not always so strongly embraced. As the evolutionary theorist Allan Orr notes in a recent review (2005), in the late nineteenth and early twentieth centuries, Mendelians like William Bateson argued that the 'micromutational' view was simply a path of least effort:

"By suggesting that the steps through which an adaptive mechanism arises are indefinite and insensible, all further trouble is spared. While it could be said that species arise by an insensible and imperceptible process of variation, there was clearly no use in tiring ourselves by trying to perceive that process. This labor-saving counsel found great favor" (Bateson 1909). The gradualist position became the dominant paradigm in the field in the 1930s via the analytical unification marrying Mendelism to Darwinism forged by R. A. Fisher, S. Wright, and J. B. S. Haldane, dubbed the "Modern Synthesis:" on this view, micro-mutational particulate events—tiny changes, perhaps at the level of single nucleotides (the DNA "letters" Adenine, Thymine, Guanine, and Cytosine), and correspondingly small steps in allele (gene variant) frequencies—comprise the bulk of evolutionary change. The Modern Synthesis too suggests that a Big-Bang emergence of language might be quite unlikely.

How then can one reconcile evolutionary theory's Modern Synthesis with the apparent discontinuity, species-specificity, and distinctive syntactic competence of human language? A familiar line of thinking simply denies that a gap exists at all: other hominids also possess human syntactic abilities. A second position embraces the Modern Synthesis and rejects discontinuity, asserting that all the particular properties of human language have been specifically selected for as directly adaptive, with small gradual changes leading from non-language using ancestors to the present, a position perhaps most strongly advocated by Pinker and Bloom (1990). Other researchers deny that there are properties proprietary to human syntax, instead grounding them on principles of general purpose cognition, like those found in connectionism (Rummelhart and McClelland 1987). Still other approaches call for development from a proto-language to full language (Bickerton 1990). Perhaps the only proposals made previously that avoid an outright appeal to gradualism are those that involve exaptation in the sense of Gould and Vrba (1982), or genetic draft (Gillespie 1991)—that human language hitchhiked on the back of a related, already adaptively advantageous cognitive subsystem, such as motor-gesture articulation, hierarchical tool-making, or social grooming. While these last approaches remain possibilities within a classical gradualist framework, they all focus on some external aspect of language-as-communication, rather than internal syntax *tout court*, a matter discussed elsewhere (Berwick and Chomsky, this volume). Berwick and Chomsky note that all recent relevant biological and evolutionary research leads to the conclusion that the process of externalization is secondary, subsequent to conceptualization and the core principles of human syntax. If this conclusion is on the right track, we are left with the same puzzling gap as before.

Besides this unsettling discontinuity, human language poses a second challenging gap, a classic and also familiar biological one: how to bridge between a genotype and a phenotype, in this case perhaps the most complex behavioral phenotype we know. Linguistics has cast natural language's intricate and ultimately behavioral "form that shows" (its phenotype) at an abstract level, far removed from language's computational and biological "inner form" or genotype. Linguistic science's successful program over the past fifty years has resulted in perhaps the richest description of a human genotype-to-phenotype mapping that we know of—the initial substrate for human language and how language develops in an individual.

However, until recently progress at bridging from the language genotype to phenotype has been difficult. In part this is because we simply do not know much about the human language genotype, and we often run aground by misidentifying the language phenotype with communication, as mentioned above. Further, the gulf separating computation, biology, and language has been equally long-standing—in large measure resulting from the abstraction gap between linguistic and biological description: we do not expect to literally find a "passive grammar rule" inside a person's head. The history of the field here from Fodor, Bever, and Garrett's summary of work from the 1960s (1974) to Berwick and Weinberg (1986), Di Sciullo (2000), Phillips (2003), Reinhart (2006), and many others might be read as one long attempt to find more-to-less isomorphic mappings between linguistic rules and representations and computational rules and representations.

In this chapter we show how to resolve both the evolutionary and the genotype–phenotype gaps in a new way, emphasizing that the species-specificity and novelty of human language need not conflict with Darwinian thinking—indeed, that modern evolutionary theorizing and discoveries have moved past the "micro-mutational" gradualist view so as to become quite compatible with modern linguistic theory, with both linguistic theory and evolutionary theory contributing theoretical insights to each other. This synergy is not a new development. It is one that itself has evolved over the past two decades: the Principles-and-Parameters (P&P) approach to language (Chomsky 1981) was directly inspired by the biologist Jacob's remarks about how the apparent diversity of biological forms might be produced by an underlying parameterization of abstract genetic regulatory switches (1977), a view that Chomsky then imported into linguistic theory. On the biological side, the so-called evo–devo [evolution–development] revolution (see Carroll, Grenier, and Weatherbee 2001; Carroll 2005; Chomsky 2007; Müller 2007), along with new results on genomic and regulatory networks, new simulation results on the apparently much larger size of adaptive mutational change

(Orr 2002, 2005), and a more widespread acknowledgment of homeoplasy or horizontal gene transfer (Warren et al. 2008), have moved evolutionary theory well past the Fisher–Wright gradualist, particulate, micro-mutational view. We shall see below that historically developments in both linguistic theory and evolutionary theory are in many respects parallel and mutually reinforcing.

To resolve these two gaps, in this chapter we move beyond the 1980s viewpoint on both the biological and linguistic fronts, showing how the most recent developments in linguistic research, dubbed the Minimalist Program (MP) (Chomsky 1995, 2007a, 2005) can bridge the biology–language divide. The MP demonstrates that despite its apparent surface complexity, language's core might in fact be much simpler than has previously been supposed. For biologists pursuing clues left by the linguistic phenotype's fault lines down to the level of the real genotype, this is a promising development. The refinement of our understanding of the linguistic phenotype comes at a particularly apt time, since in the last decade there has been an ever-growing, though still small, range of work on genetics and language, as exemplified in the work of Gopnik and colleagues (1990), and many others since, including the extensive recent findings on locating specific genetic variation in language, namely, the mutations in the FOXP2 gene (Marcus and Fisher 2003; Enard et al. 2005). (But see Berwick and Chomsky, this volume, for a critique of naive genetical interpretations of single mutations.) This is because the minimalist program is eliminative in exactly the right way, and so can serve as a case study for how a complex behavioral phenotype emerges from the interactions of a much simpler genotype. In particular, the minimalist program posits that the human syntactic engine consists of just two components: (1) words and word features; and (2) a single, simple recursive operation, Merge, that glues together words and word complexes into larger units.

This chapter demonstrates how *just* these two components, without further stipulation, interact to yield many, perhaps all, of the special design features of human language syntax. If this is so, then we have no need for specific, adaptive accounts of these particular features. By design features we mean familiar properties of human language syntax such as the following:

- digital infinity and recursive generative capacity, the familiar 'infinite use of finite means:' sentences may be arbitrarily long and novel; there are 1-, 2-, . . . word sentences, but there are no $5\frac{1}{2}$-word sentences;
- displacement: human languages move phrases from their natural argument positions, as in *This student, I want to solve the problem* where the subject

of the verb, *solve*, namely *This student*, appears at the front of the sentence instead of in its normal position after the verb;

- locality constraints: displacement does not act over unbounded domains—in *Who do you wonder Bill thinks solved the problem*, *who* cannot be interpreted as the subject of *solve*;
- restricted grammatical relations: out of a potentially infinite set of logically possible relations that might be defined over configurations of syntactic structures, only a handful ever seem to play a role in human syntax. For example, human languages often match verbs to objects (in terms of predicate–argument structure); demand agreement between tense/inflection and subjects as in the case of subject–verb person–number agreement; or verbs may select either subjects or objects, as in the familiar contrast between *John admires honesty*/*Honesty admires John*. Yet most logically possible syntactic rules and relations are unattested—for instance, there is apparently no analog to 'object-of,' say subject-object-of, where the subject and object of a sentence must agree.

For the evolutionary biologist seeking to answer *why* we see this particular distribution of organisms or traits in the natural world and not others—one of the central questions of biology being to reconcile this pattern of variation both present and absent with the apparent commonalities of organisms, just as with language—such a finding is central. If observed patterns follow from a single, central principle, then there is no need to invoke some special adaptive explanation for any of them. There is no locality "trait" and no grammatical relation trait that must be acquired in an evolutionary piecemeal fashion. One does not need to advance incremental, adaptationist arguments with intermediate steps between some protolanguage and full natural language to explain much, perhaps all, of natural language's specific design.

Note that from a logical or communicative standpoint, these particular design properties are otherwise mysterious. For instance, there is no immediately obvious computational or communicative reason why languages ought *not* to relate subjects and objects. Communicatively, a sentence's subject, usually an agent, and its object, usually the affected object, form just as natural a class as subject and predicate. Further, as is easy to see from the transitivity of conditional probabilities that can be simply multiplied together, nothing blocks a purely statistical conditional relationship between subject and object. However, it seems that no such connections are to be found in human languages. Indeed this is another clear limitation of the currently popular statistical approach to language description, which otherwise offers no barrier to such unattested relations. Similarly, as pointed out in Berwick and Chomsky (this volume), displacement makes language processing and

communication *more* difficult, not *less* difficult—yet another argument that language is not designed for communication. The ultimate explanation for language's design must be, obviously, biological, but on the view here, not at the level of expressiveness or communicative efficiency. This chapter offers an alternative, deeper, possibility: the reason why human syntax looks the way it does rather than some other way—why natural languages have an object-of grammatical relation but not a subject-object-of grammatical relation—*follows* from the fundamental principles of the basic combinatorial syntactic engine itself.[1]

As to the evolutionary origin of the fundamental combinatory ability itself, Merge, we leave this topic largely unexplored here. Along with the evolutionary theorist G. C. Williams (1996: 77), one might speculate that the hierarchical combinatorial ability possibly appeared just as other evolutionary novelties do: "new structures arise in evolution in one of two ultimate ways, as redundancies or *spandrels*"—a structure arising as an incidental consequence of some other evolutionary change. Where we part ways with Williams's classical account is in the nature of evolutionary change itself, as we describe in Section 4.2, below. Berwick and Chomsky (this volume) provide additional details on how a singular event of this kind might arise and spread in a small group, some 50,000 to 100,000 years ago.

This minimalist reformulation also has important consequences for models of language processing, and so ultimately for descriptions of the linguistic phenotype as it is externalized. The most minimal conception of a processor or parser for natural language takes the relation between basic parsing operations and the abstract linguistic system as simply the identity function. As it turns out, this leads to the most efficient processor achievable, left-to-right,

[1] We note that it is more difficult than sometimes supposed to give a purely functional communicative justification for some of the more patent universal properties of natural-language syntax. For instance, it is sometimes suggested that recursive generative capacity is somehow *necessary* for communication, thus bridging the gap between protolanguage and recursive human syntax. But is this so? There seem to be existing human languages that evidently possess the ability to form recursive sentences but that apparently do not need to make use of such power: a well-known example is the Australian language Warlpiri, where it has been proposed that a sentence that would be recursively structured in many other languages, such as 'I think that John is a fool' is formed via linear concatenation, 'I ponder it. John is a fool'; or to take a cited example, *Yi-rna wita yirripura jaru jukurrpa-warnu wiinyiinypa*, literally, 'little tell-PRESENT TENSE story dreaming hawk,' translated as 'I want to tell a little dreaming story about a hawk' (Nash 1986, Swartz 1988). Evidently then, recursion is not essential to express "the beliefs about the intentional states of others," quite contrary to what some researchers such as Pinker and Bloom (1990) and more recently Pinker and Jackendoff (2005) have asserted. This seems again to be an example of the confusion between externalization-as-communication and internal syntax. Apparently the same was true of Old English, if the data and linguistic arguments presented in O'Neil (1976) are correct. There is nothing surprising in any of this; it is quite similar in spirit to the observed production/perception limitations on, for example, center-embedded and other difficult-to-process sentences in English.

real-time, and deterministic to the extent this is possible at all, and at the same time replicates some of human language's known psychophysical, preferential 'blind spots.' For example, in sentence pairs such as *John said that the cat died yesterday*/*John said that the cat will die yesterday*, *yesterday* is (reflexively) taken to modify the second verb, the time of the cat's demise, even though this is semantically defeasible in the second sentence. In this sense, this approach even helps solve the secondary process of externalization (and so communicative efficiency, again to the extent that efficient communication is possible at all).

If all this is correct, then current linguistic theory may have now attained a better level of description in order to proceed with evolutionary analysis. In this sense, using familiar Darwinian terms, the syntactic system for human language is indeed, like the eye, an "organ of extreme complexity and perfection." However, unlike Linnaeus's and Darwin's slogan shunning the possibility of discontinuous leaps in species and evolution generally—*natura non facit saltum*—we advocate a revised motto that turns the original on its head: *syntax facit saltum*—syntax makes leaps—in this case, because human language's syntactic phenotype follows from interactions amongst its deeper components, giving it a special character all its own, apparently unique in the biological world.

The remainder of this chapter is organized as follows. Section 2 serves as a brief historical review of the recent shift from the "gradualist," micro-mutational Modern Synthesis in evolutionary biology to a more ecumenical encompassing the evo–devo revolution and macro-adaptive events. Section 3 follows with an outline of parallel shifts in linguistic theory: from highly specialized language-particular rules to more abstract principles that derive large-scale changes in the apparent surface form of syntactic rules. This historical shift has two parts. First, the change from the atomic, language-particular rules of the *Aspects* era, the Extended Standard Theory (EST) to a system resembling a genetic regulatory–developmental network, dubbed the Principles-and-Parameters (P&P) theory; and second, the more recent shift to the Minimalist Program. Putting to one side many important details irrelevant for our argument, we outline how sentence derivations work in the MP. By examining the notion of derivation in this system, we demonstrate that all syntactic 'design features' in our list above follow ineluctably, including all and only the attested grammatical relations, such as subject-of and object-of. Section 4.4 turns to sentence processing and psychophysical blind spots. It outlines a specific parsing model for the minimalist system, based on earlier computational models for processing sentences deterministically, strictly left to right. It then shows how reflexive processing preferences like the one

described above can be accounted for. Section 4.5 concludes with observations on the tension between variation and uniformity in language, summarizing the evolutionary leap to syntax.

4.2 Ever since Darwin: The Rise and Fall of Atomism in Modern Evolutionary Biology

As we noted in the introduction, by the mid-1930s the micro-mutational, atomistic picture of evolution by natural selection held sway, admirably unified on the one hand by particulate Mendelism and on the other hand by Darwinism, all unified by the 'infinitesimal' mathematical theory established by Fisher, Wright, and Haldane. However, by the late decades of the twentieth century and on into the twenty-first, this Modern Synthesis paradigm has undergone substantial revision, due to advances on three fronts: (1) the evo–devo revolution in our understanding of deep homologies in development and underlying uniformity of all organisms; (2) an acknowledgment of a more widespread occurrence of symbiotic evolutionary events and homeoplasy; and (3) the erosion of the Fisher-inspired adaptationism-as-incrementalism model.

First, the evo-devo revolution has demonstrated that quite radical changes can arise from very slight changes in genomic/developmental systems. The evolution of eyes provides a now-classic example, turning Darwin's view on its head. While perhaps the most famous advocate of the Modern Synthesis, Ernst Mayr, held that it was quite remarkable that 40 to 60 different eyes had evolved separately, thus apparently confirming that micro-mutational evolution by natural selection had a stunning ability to attain 'organs of perfection' despite radically different starting points and different contexts, in reality there are very few types of eye, perhaps even monophyletic (evolving only once) 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 (Salvini-Plawen and Mayr 1961; Gehring 2005). Indeed, this example almost exactly parallels the discoveries in linguistic theory during the 1980s, that the apparent surface variation among languages as distinct as Japanese, English, Italian, and so forth are all to be accounted for by a richly interacting set of principles, parameterized in just a few ways, that deductively interact to yield the apparent diversity of surface linguistic forms, as described in the next section.

This move from a highly linear model—compatible with a 'gradualist, incrementalist' view—to a more highly interconnected set of principles where a slight change in a deep regulatory switch can lead to a quite radical sur-

face change, from vertebrate eyes to insect eyes—follows the same logic in both linguistics and biology. The parallel is much more than superficial. Gehring (2005) notes that there is a contrast to be made in the evolution of biosynthetic pathways, for instance for histidine, as proposed by Horowitz (1945), as opposed to the evolution of morphogenetic pathways, as for the vertebrate eye. In the case of biosynthesis, the linear model can proceed backwards: at first, histidine must be directly absorbed from the environment. When the supply of histidine is exhausted, then those organisms possessing an enzyme that can carry out the very last step in the pathway to synthesize histidine from its immediate precursor are the Darwinian survivors. This process extends backwards, step by step, linearly and incrementally. We might profitably compare this approach to the one-rule-at-a-time acquisition that was adopted in the Wexler and Culicover model of acquisition in the similarly atomistic linguistic theory of the time (1980).

However, quite a different model seems to be required for eye morphogenesis, one that goes well beyond the incremental, single nucleotide changes invoked by the Modern Synthesis. Here the interplay of genetic and regulatory factors seems to be intercalated, highly interwoven as Gehring remarks. That much already goes well beyond a strictly micromutational, incremental view. The intercalation resembles nothing so much as the logical dependencies in the linguistic P&P theory, as illustrated in Figure 4.1 see p. [ref to fo. 149]. But there is more. Gehring argues that the original novelty itself—a photoreceptor next to a pigment cell—was a "purely stochastic event," and further, that 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. Needless to say, this completely bypasses the ordinary step-by-step nucleotide changes invoked by the Fisher model, and constitutes the second major discovery that has required re-thinking of the gradualist, atomistic view of evolution:

Au: I believe this is p. 82

> The eye prototype, which is due to a purely stochastic event that assembles a photoreceptor and a pigment cell into a visual organ, requires the function of at least two classes of genes, a master control gene, *Pax6*, and the structural genes encoding on rhodopsin, for instance, the top and the bottom of the genetic cascade. Starting from such a prototype increasingly more sophisticated eye types arose by recruiting additional genes *into* the morphogenetic pathway. At least two mechanisms of recruitment are known that lead to the intercalation of additional genes into the genetic cascade. These mechanisms are gene duplication and enhancer fusion. . . . For the origin of metazoan photoreceptor cells I have put forward two hypotheses: one based on cell differentiation and a more speculative model based on symbiosis.
>
> (2005: 180; emphasis added)

The outcome is a richly interconnected set of regulatory elements, just as in the P&P theory. Further, we should note that if the origin of the original novelty assembling photoreceptor and pigment cell is purely stochastic then the "cosmic ray" theory of the origin of Merge, sometimes derided, might deserve more serious consideration, though of course such a view must remain entirely speculative.

Second, the leaps enabled by symbiosis seem much more widespread than has been previously appreciated. As Lynn Margulis—the biologist who did much to establish the finding that cellular mitochondria with their own genes were once independent organisms ingested symbiotically—has often remarked, "the fastest way to get new genes is to eat them." To cite yet another recent example here out of many that are quickly accumulating as wholesale genomic analysis accumulates, the sequencing of the duckbill platypus *Ornithorhynchus anatinus* genome has revealed a substantial number of such horizontal transfers of genes from birds and other species whose most recent common ancestor with the platypus is extremely ancient (Warren et al. 2008). (Perhaps much to the biologists' relief, the crossover from birds does not seem to include the genes involved in the platypus's duck bill.)

Of course at the lowest level, by and large genomic changes must of necessity be particulate in a strong sense: either one DNA letter, one nucleotide, changes or it does not; but the by-and-large caveat has loomed ever larger in importance as more and more evidence accumulates for the transmission of genes from species to species horizontally, presumably by phagocytosis, mosquito-borne viral transmission, or similar processes, without direct selection and the vertical transmission that Darwin insisted upon.

The third biological advance of the past several decades that has eroded the micro-mutational worldview is more theoretical in character: Fisher's original mathematical arguments for fine-grained adaptive change have required substantial revision. Why is this so? If evolutionary hills have gentle slopes, then inching uphill always works. That follows Fisher chapter and verse: picture each gene that contributes to better eyesight as if it were one of millions upon millions of fine sand grains. Piling up all that sand automatically produces a neatly conical sand pile with just one peak, a smooth mound to climb. In this way, complex adaptations such as the eye can always come about via a sequence of extremely small, additive changes to their individual parts, each change selectively advantageous and so seized on by natural selection.

The key question is whether the biological world really works this way, or rather, how often it works this way. And that question divides into two parts. Theoretically speaking: what works better as the raw material or "step size"

for adaptation—countless genes each contributing a tiny effect, or a handful of genes of intermediate or large effect? Empirically speaking: how does adaptation really play out in the biological world? Are large mutations really always harmful, as Fisher argued? Do organisms usually tiptoe in the adaptive landscape or take larger strides? Are adaptive landscapes usually smooth sand piles, jagged alpine ranges, or something in between?

Fisher addressed the theoretical question via a mathematical version of the familiar "monkey wrench" argument: a large mutation would be much more likely than a small one to gum up the works of a complex, finely constructed instrument like a microscope, much as a monkey randomly fiddling with the buttons on a computer might likely break it. It is not hard to see why. Once one is at a mountain top, a large step is much more likely to lead to free-fall disaster. But the microscope analogy can easily mislead. Fisher's example considers a mutation's potential benefits in a particularly simple setting—precisely where there is just one mountain top, and in an infinite population. But if one is astride K90 with Mt. Everest just off to the left, then a large step might do better to carry me towards the higher peak than a small one. The more an adaptive landscape resembles the Himalayas, with peaks crowded together—a likely consequence of developmental interactions, which crumple the adaptive landscape—the worse for Fisher's analogy. Small wonder then that Dawkins's topographic maps and the gradual evolutionary computer simulations he invokes constantly alter how mountain heights get measured, resorting to a single factor—first for eyes, it's visual resolution; next, for spider webs, it is insect-trapping effectiveness; then, for insect wings, it is aerodynamic lift or temperature-regulating ability. An appropriate move, since hill-climbing is guaranteed to work only if there is exactly one peak and one proxy for fitness that can be optimized, one dimension at a time.

Even assuming a single adaptive peak, Fisher's microscope analogy focused on only half the evolutionary equation—variation in individuals, essentially the jet fuel that evolution burns—and not the other half—the selective engine that sifts variations and determines which remain written in the book of life. Some fifty years after Fisher, the population biologist Motoo Kimura (1983) noted that most *single* mutations of small effect do not last: because small changes are only slightly selectively advantageous, they tend to peter out within a few generations (ten or so). Indeed, most mutations, great or small, advantageous or not, go extinct—a fact often brushed aside by pan-adaptationist enthusiasts (see also Berwick and Chomsky, this volume). Kimura calculated that the rate at which a mutation gains a foothold and

then sweeps through a population is directly proportional to the joint effect of the probability that the mutation is advantageous and the mutation's size. Moreover, even if medium-scale changes were less likely to fix in a population than micro-mutations, by definition a larger change will contribute correspondingly more to an organism's overall response to natural selection than a small one.

However, Kimura neglected an important point: he calculated this relative gain or loss for just a single mutation, but the acquisition of some (perhaps complex) trait might take several steps. This alters Kimura's analytical results, as has been more recently studied and thoroughly analyzed by Orr via a combination of mathematical analysis and computational simulations (2002, 2005), and extended to discrete molecular change in DNA sequence space via other methods by Gillespie (1984). The upshot seems to be that beneficial mutations have exponentially distributed fitness effects—that is,

> adaptation is therefore characterized by a pattern of diminishing returns—larger-effect mutations are typically substituted earlier on and smaller-effect ones later, . . . indeed, adaptation seems to be characterized by a 'Pareto principle', in which the majority of an effect (increased fitness) is due to a minority of causes (one [nucleotide] substitution).
> (Orr 2005: 122, 125)

Thus, while Kimura did not get the entire story correct, Fisher's theory must be revised to accommodate the theoretical result that the first adaptive step will likely be the largest in effect, with many, many tiny steps coming afterwards. Evidently, adaptive evolution takes much larger strides than had been thought. How then does that connect to linguistic theory? In the next section, we shall see that linguistics followed a somewhat parallel course, abandoning incrementalism and atomistic fine-grained rules, sometimes for the same underlying conceptual reasons.

4.3 Ever since *Aspects*: The Rise and Fall of Incrementalism in Linguistic Theory

Over the past fifty years, linguistic science has moved steadily from less abstract, naturalistic surface descriptions to more abstract, deeper descriptions—rule systems, or generative grammars. The Minimalist Program can be regarded as the logical endpoint of this evolutionary trajectory. While the need to move away from mere sentence memorization seems clear, the rules that linguists have proposed have sometimes seemed, at least to some, even farther removed from biology or behavior than the sentences they were

meant to replace. Given our reductionist aim, it is relevant to understand how the Minimalist Program arose out of historical developments of the field, partly as a drive towards a descriptive level even farther removed from surface behavior. We therefore begin with a brief review of this historical evolution, dividing this history into two parts: from the Extended Standard Theory of Chomsky's *Aspects of the Theory of Syntax* (1965) to the P&P model; and from P&P to the Minimalist Program.

Before setting out, it perhaps should first be noted that the 'abstraction problem' described in the introduction is not unfamiliar territory to biologists. We might compare the formal computations of generative grammar to Mendel's Laws as understood around 1900—abstract computations whose physical basis were but dimly understood, yet clearly tied to biology. In this context one might do well to recall Beadle's comments about Mendel, as cited by Jenkins (2000):

> There was no evidence for Mendel's hypothesis other than his computations and his wildly unconventional application of algebra to botany, which made it difficult for his listeners to understand that these computations *were* the evidence.
>
> (Beadle and Beadle 1966)

In fact, as we suggest here, the a-biological and a-computational character sometimes (rightly) attributed to generative grammar resulted not because its rules were abstract, but rather because rules were not abstract enough. Indeed, this very fact was duly noted by the leading psycholinguistic text of that day: Fodor, Bever, and Garrett's *Psychology of Language* (1974: 368), which summarized the state of psycholinguistic play up to about 1970: "there exist no suggestions for how a generative grammar might be concretely employed as a sentence recognizer in a psychologically plausible system."

In retrospect, the reason for this dilemma seems clear. In the initial decade or two of investigation in the era of modern generative grammar, linguistic knowledge was formulated as a large set of language-particular, specific rules, such as the rules of English question formation, passive formation, or topicalization (the rule that fronts a focused phrase, as in, *these students I want to solve the problem*). Such rules are still quite close to the external, observable behavior—sentences—they were meant to abstract away from.

4.3.1 *All transformations great and small: From Aspects to Principles and Parameters*

By 1965, the time of Chomsky's *Aspects of the Theory of Syntax*, each transformational rule consisted of two components: a structural description, generally corresponding to a surface-oriented pattern description of the conditions

under which a particular rule could apply (an 'IF' condition), and a structural change marking out how the rule affected the syntactic structure under construction (a 'THEN' action). For example, an English passive rule might be formulated as follows, mapping *Sue will eat the ice cream* into *The ice cream will be+en eat by Sue*, where we have distinguished pattern-matched elements with numbers beneath:

Structural description (IF condition):

Noun phrase	Auxiliary Verb	Main Verb	Noun Phrase
1	2	3	4
Sue	will	eat	the ice cream

Structural change (THEN):

Noun phrase	Auxiliary Verb	be+en Main	Verb by	Noun Phrase
4	2		3	1
The ice cream	will	be+en	eat	by Sue

In the *Aspects* model a further housekeeping rule would next apply, hopping the *en* affix onto *eat* to form *eaten.*

This somewhat belabored passive-rule example underscores the non-reductionist and atomistic, particulate, flavor of earlier transformational generative grammar: the type and grain size of structural descriptions and changes do not mesh well with the known biological descriptions of, for example, observable language breakdowns. Disruption does not seem to occur at the level of individual transformational rules, nor even as structural descriptions and changes gone awry generally.

The same seems to holds true for many of the biological/psycholinguistic interpretations of such an approach: as Fodor, Bever, and Garrett remarked, individual transformational rules did not seem to be engaged in sentence processing, and the same problems emerge when considering language learnability, development, or evolution. Indeed, the biological/evolutionary picture emerging from the *Aspects*-type (Extended Standard Theory or EST) grammars naturally reflects this representational granularity. For any given language, in the EST framework, a grammar would typically consist of many dozens, even hundreds, of ordered, language-specific transformations, along with a set of constraints on transformations, as developed by Ross (1967), among others. This 'particulate' character of such rules was quite in keeping with a micro-mutational view: if the atoms of grammar are the bits and pieces of individual transformations, structural descriptions and changes, then it is quite natural to embrace a concomitant granularity for thinking about language learning, language change, and language

evolution. It is quite another matter as to the empirical reality of this granularity.

Consider by way of example the Wexler and Culicover model of the late 1970s, establishing that EST grammars are learnable from simple positive example sentences (1980). In the Wexler and Culicover model, what was acquired at each step in the learning iteration was a single rule with a highly specific, structural description/structural change, as driven by an error-detection principle. This atomistic behavior was quite natural, since it mirrored the granularity of the EST theory, namely, some sequence of transformations mapping from a base D-structure to a surface S-structure.

But equally, this extreme particulate view was embraced in other contexts, notably by Pinker and Bloom (1990), who argued that language evolution must follow a similar course: *all* of the specific design features of language *must* have arisen by incremental, adaptive evolutionary change. What Pinker and Bloom did in part is to run together two distinct levels of representation: what of necessity must be true, barring horizontal transfer—that genomic evolution lies, ultimately, at the level of single DNA letters or nucleotides—with what need not, and we shall argue, is not true—that language evolution, a distal behavior, proceeds apace at the level of the granularity set by the EST linguistic theory, with all evolutionary change incremental and adaptive. Such a position can lead to difficulties. It argues in effect that each and every property of syntactic design that we see must have some measurable outcome on fitness—perhaps quite literally, the number of viable offspring. As Lightfoot (1991) notes, this entails what might be dubbed the "Subjacency and Sex Problem"—there are important syntactic constraints, like the one called "subjacency," a restriction on the distance that a phrase can be displaced—that have no obvious effect on the number of offspring one might have, absent special pleading: "subjacency has many virtues, but . . . it could not have increased the chances of having fruitful sex." Further, this stance seemingly entails connecting all such "inner" constraints once again somehow to external communication—a delicate, probably incorrect link, as we have seen. Indeed, Pinker and Bloom do not advance any specific evolutionary modeling details at all about which syntactic structures are to be linked to particular aspects of fitness so as to construct a proper evolutionary model. Fortunately, we do not have to resort to this level of detail. As we have outlined in the previous section, such incrementalism and pan-adaptationism is now known to be far from secure quite generally in evolutionary biology, even assuming the conventional Fisher model. If one in addition takes into account our more recent, nuanced understanding of the rapid evolution that can occur due to developmental changes as well as the widespread possibility of

978–0–19–955327–3 04-Sciullo-c04 Di-Sciullo (Typeset by SPi, Chennai) 81 of 99 May 10, 2010 13:46
OUP UNCORRECTED PROOF – FIRST PROOF, 10/5/2010, SPi

homeoplasy, then this micro-mutational view of language evolution fares even worse.

Given such rule diversity and complexity, by the mid-1960s the quasi-biological problems with surface-oriented rules—problems of learnability and parsability among others—were well known: how could such particular structural conditions and changes be learned by children, since the evidence that linguists used to induce them was so hard to come by? The lack of rule restrictiveness led to attempts to generalize over rules, for example, to bring under a single umbrella such diverse phenomena as topicalization and question formation, each as instances of a single, more general, Move *wh*-phrase operation. By combining this abstraction with the rule Move Noun phrase, by the end of the 1970s linguists had arrived at a replacement for nearly all structural changes or displacements, a single movement operation dubbed Move alpha. On the rule application side, corresponding attempts were made to establish generalizations about constraints on rule application, thereby replacing structural descriptions—for example, that noun phrases could be displaced only to positions where they might have appeared anyway as argument to predicates, as in our passive example.

Inspired in part by a lecture given by Jacob at MIT's Endicott House in 1974 on the topic of biological variation as determined by a system of genetic switches (see Berwick and Chomsky, this volume), Chomsky had already begun to attempt to unify this system of constraints along the same lines—a concrete example of linguistics inspired by biology. By the 1980s, the end result was a system of approximately 25 to 30 interacting principles, the so-called Principles-and-Parameters (P&P) or Government-and-Binding approach (Chomsky 1981). Figure 4.1 sketches P&P's general picture of sentence formation, shaped like an inverted Y. This model engages two additional representational levels to generate sentences: first, D-structure, a canonical way to represent predicate–argument thematic relations and basic sentence forms—essentially, "who did what to whom," as in *the guy ate the ice cream* where *the guy* is the consumer and *ice cream* is the item consumed; and second, S-structure, essentially a way to represent argument relations after displacement—like the movement of the object to subject position in the former passive rule—has taken place. After the application of transformations (movement), S-structure splits, feeding sound (phonological form, PF) and logical form (LF) representations to yield (sound, meaning) pairs.

Overall then, on the Principles-and-Parameters view, sentences are derived beginning with a canonical thematic representation that conforms to the basic tree structure for a particular language, and then mapped to S-structure via

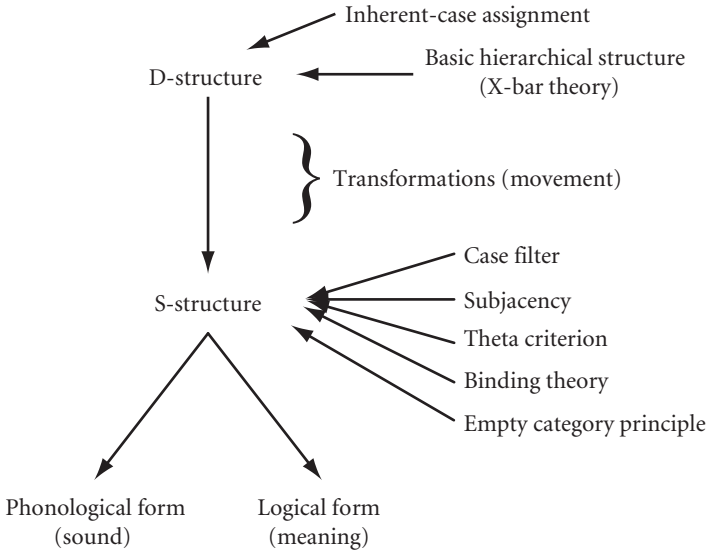

FIGURE 4.1 A conceptual picture of the traditional transformational generative-grammar framework (applying equally to the Extended Standard Theory, Government–Binding, and Principles-and-Parameters, approaches). Thin arrows denote constraints that possible sentence forms must satisfy, like the case filter. We do not describe all the depicted constraints in this chapter.

a (possibly empty) sequence of displacement operations. For instance, one could start with *the guy ate the ice cream* in hierarchical form with a thematic or D-structure; via displacement of *the ice cream* this initial representation can be mapped to the topicalized form, *ice cream, the guy ate*. To use another conceptual picture that has been developed for computer implementation, in the P&P approach, sentence generation can be viewed as starting at D-structure and then running a gauntlet through a set of constraint boxes placed at D-structure and S-structure, as shown in Figure 4.1. A sentence is completely well formed if it passes all the constraints and emerges at the two interfaces of phonological form and logical form as one or more sound, meaning pairs.

Akin to atomic theory, this small set of constraints may be recombined in different ways to yield the distinctive syntactic properties of diverse natural languages, just as a handful of elements recombine to yield many different molecular types, or, anticipating the evo–devo revolution, the way that the same regulatory genes might readjust their timing and combinations to yield "endless forms most beautiful" in Darwin's famous closing lines, the title of Carroll's recent book on evo–devo (2005). For example, one of

the principles, X-bar theory, constrains the basic D-structure tree shapes for phrases—whether phrases can appear in function–argument form, as in English verb–object or preposition–object combinations, for example, *eat ice cream* or *with a spoon*, or alternatively, in argument–function form, as in Japanese object–verb or postposition–object combinations, such as *ice cream-o tabeta* or *spoon-ni*.

The X-bar module constrains just a small part of the ultimate surface form of sentences and must conspire with other principles to yield the surface complexity that one actually sees. In order to replicate the passive rule, at least three other general principles constraining displacement and S-structure come into play. One such constraint is the so-called theta criterion: if one pictures a verb as a predicate taking some number of arguments—its thematic roles, such as *drink* requiring something to be drunk—then at the end of a derivation, all of the verb's arguments must have been discharged or realized in the sentence, and every possible argument in the sentence must have received some thematic role. A second constraint is the Case Filter: any pronounceable noun phrase, such as *the guy*, must possess a special feature dubbed Case, assigned by a verb, preposition, or tense/inflection.

Now the former passive rule follows as a theorem from these more basic principles. Starting from the D-structure *was eaten ice cream*, since *eaten* does not assign Case (analogously to an adjectival form, like *tired* or *happy*), the object *ice cream* must move to a position where it does get case—namely, the position of the subject, where *ice cream* can receive case from the inflected verb *was*. We thus derive the surface form *the ice cream was eaten*. The thematic association between *eat* and *ice cream* as the material eaten is retained by a bit of representational machinery: we insert a phonologically empty (unpronounced) element, a trace, into the position left behind by *ice cream* and link it to *ice cream* as well. In a similar fashion one can show that approximately thirty such constraints suffice to replace much of syntax's formerly rule-based core. The end result is a system very close to Jacob's system of genetic regulatory "switches," with variation among languages restricted to the choice of parameters such as function–argument or argument–function form, possibly restricted further to choices of lexical variation. The language phenotype now looks rather different; it has far less to do with the surface appearance of structure descriptions and structural change, but is constituted entirely of the parameters and their range of variation, a rather different picture. This P&P approach to language variation was quite fruitful, and led to models of language acquisition, change, and parsing that differed substantially from the *Aspects* view, more closely mirroring the possibilities of radical surface differences given just a few underlying changes. For example,

in the domain of language acquisition and change, Niyogi and Berwick (1996) demonstrated the possibility of rapid phase changes from one language type to another, for example, from the verb-final form of earlier English to modern-day English. In the domain of parsing, Fong (1990) implemented a uniform computational engine with twenty-odd modular components, corresponding to the Case filter, thematic role checking, X-bar theory, and so forth, parameterized along narrow lines like genetic switches to yield a unified parser for English, Japanese, German, Hungarian, Turkish, and many other languages. None of these possibilities were realizable under the *Aspects* model.

4.3.2 *From P&P to the Minimalist Program: Reducing the language phenotype*

The Minimalist Program goes the Principles-and-Parameters approach one better: it aims to eliminate all *representations* and *relations* that can be derived from more primitive notions. Syntax still mediates form and meaning in the classical Saussurean sense, as in Figure 4.1 with its paired sound and meaning interfaces—but the representations of D-structure and S-structure are eliminated. To build syntactic objects and relations, minimalism invokes only the notion of a word construed as a list of features plus a generalized hierarchical derivation operator, called Merge. For example, it is Merge that glues together *eat* and *ice cream* to form the verb phrase *eat ice cream* and tacks the *en* morpheme onto the end of *eat* to form *eaten*; a sequence of Merges generates a sentence. In fact, relationships among syntactic objects established by Merge constitute the totality of syntactic structure, and, as we shall see, also fix the range of syntactic relations. In other words, those elements that enter into the Merge operation are precisely those that can be syntactically related to each other. Merge thus delimits the atoms and molecules visible for chemical combination. At the sound–meaning interfaces, the only available entities are syntactic objects and the syntactic structures these objects form. These entities contain inherent word features that impose constraints on articulatory generation or parsing and conceptual–intentional interpretation. What drives the generative process is feature matching and feature elimination, as we now describe.

4.3.2.1 *Deriving sentences in the minimalist approach* To see how this generative machinery works, let us consider a concrete example that we will follow through the remainder of this chapter. The following two figures illustrate, with Figure 4.2 providing a conceptual overview and Figure 4.3 more detail. We retain the basic syntactic categories from previous syntactic models, considered as features, both open-class categories such as n(oun) and v(erb), as well as grammatical categories like d(eterminer), t(ense) (or i(nflection)),

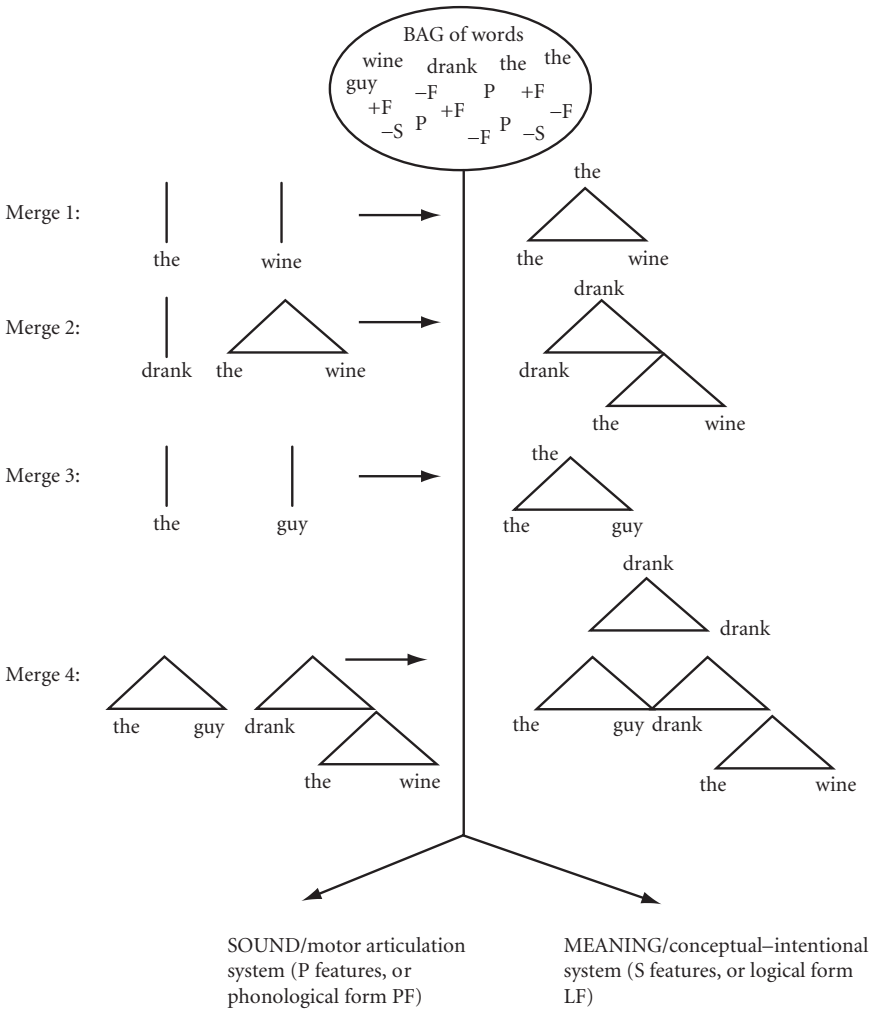

FIGURE 4.2 Merge maps from an initial array of words to a (sound, meaning) pair—representations containing only phonological or semantic features, respectively. A sequence of Merge operations constitutes a derivation in the Minimalist Program, generating a sentence from an initial unordered word set. *P*, *S*, and *F* denote phonological, semantic, and formal (syntactic) features, respectively

c(omplementizer), and so forth. Conceptually, in Figure 3.2 we begin with an unordered bag of words (formally, a multiset, since some words may be repeated), where words are just feature bundles as we describe in more detail later. In our example, we begin with {the, guy, drank, the, wine} and via four derivational steps, four Merges, wind up with the syntactic structure

4
Au: Please confirm the figure no.

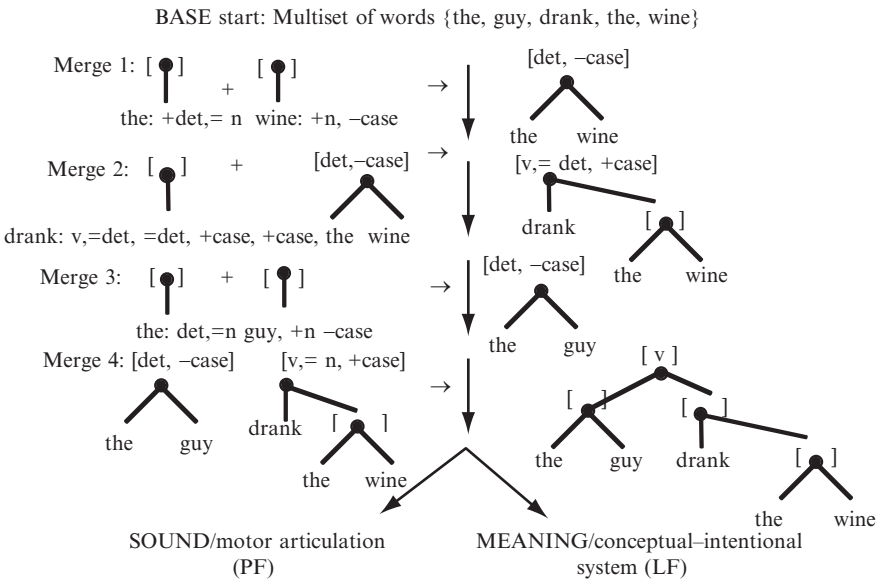

FIGURE 4.3 Details of sentence derivation in the minimalist system. The basic derivational operator, Merge, applies four times, starting with an unordered word multiset. Each Merge combines either two words, indicated by straight lines, into a hierarchical superword, indicated by a triangle, or else combines two word/hierarchical superwords into a new hierarchical combination. Merger continues until all possible formal features have been eliminated

corresponding to *the guy drank the wine*, which is then spun off to both phonological and semantic interpretation. We should emphasize at the outset that these figures depict just one possible, successful, derivational sequence. In fact, with five words there are 5! or 120 possible basic derivational possibilities, but most of these, as we shall see, do not lead to well-formed structures.

Adhering to the chemical analogy of sentence derivation, minimalism deploys Merge to combine words into larger, hierarchical superwords via a notion somewhat like chemical valency. All structure-building is feature driven, via words with formal (F), phonological/sound (P), and semantic (S) features. Figure 4.2 depicts these as F, P, and S features in an initially unordered word soup (a lexicon). Roughly, phonological features are those that can be interpreted, or 'read' by the articulatory/perceptual interface—such as classical distinctive features that refer to articulation, like ± Coronal; semantic features are those that can be read by the conceptual/intentional interface—such as ± Past; while formal (or syntactic) features include those such

as ± Tns or ± Case that play no role in sound or meaning. Syntactic features also encompass the traditional notion of *selection*, as in the sense of agreement or a verb–argument relation: a feature attached to a word can select for a particular syntactic category feature to its right or left. Following Stabler (1997) we use the notation = x to denote such a requirement; for example, the feature = n means that a verb like *drink* could select a word marked with the feature n (for noun) to its right or left. (Again, a more careful approach like that advanced in Di Sciullo 2005 ought to be able to derive this asymmetrical selection from asymmetrical Merge itself.)

Merge combines two words, a word plus an affix, or two word complexes into a single new superword, with one of the two elements being selected as the head of the new hierarchical structure, as shown schematically as the combination of two vertical lines into a triangle or two triangles into a larger one. Merge is triggered in two ways: (1) either $+f$ and $-f$ formal features on two words or word complexes can cancel, erasing the formal feature f; or (2) $\mathrm{a} = x$ feature can select $\mathrm{a} + x$ category. For example, we take *the* to be a determiner, $+det$, selecting the feature n(oun), so it has the formal features $-det$, $=n$; while *wine* is marked $+n$, $-Case$. The $= n$ feature can select the $+n$ feature, so Merge can apply in this case. (Right/left order is irrelevant for selection; we put to one side the important question of how actual word order is fixed, for instance., why the combination *wine the* is barred.) Merging these two words,*the* is taken as the head of a new hierarchical complex, which one can write as {the {*the, wine*}}, and which would traditionally have been written as a phrase-structure tree. The process of matching and canceling features, or matching and selecting features, is called *feature checking*. Note that it is possible for Merge to fail if features do not cancel or match: for instance, we cannot Merge *wine* and *guy*. Finally, it is important to add that Merge is driven by a locality notion of economy: a feature $-f$ must be checked done "as soon as possible"—that is, by the closest possible corresponding $+f$ feature. (For a much broader and more sophisticated view of selection and feature checking that can derive some of these properties, driven by the asymmetrical nature of syntactic relations generally, including morphology, see Di Sciullo 2005).

After Merge applies, any features that remain unchecked are copied or projected to the top of the new hierarchical structure, so our example complex has the features $+det$, $-Case$; conventionally, a noun phrase. (We could just as easily envision this as copying the entire word *the* to the head of the new structure, as shown in Figure 2.) Note that on this view, it is only words and affixes—the leaves of a syntactic structure—that have features; the head of a hierarchical structure receives its features only from these.

Merge operations repeat until no more features can be canceled, as shown in Figure 4.2, and in detail in Figure 4.3—note that after step 4, all formal syntactic features have been eliminated and only sound and meaning features remain to be read by the phonological and conceptual–intentional machinery, a process dubbed spell-out. Note that in fact spell-out is possible at any time, so long as the structure shipped off to PF or LF is well formed.

Step by step, generation proceeds as follows (see Figure 3). Selecting a possible Merge at random, *the* $\{+det, =n\}$ can combine with *wine* $\{+n, -case\}$, selecting the $+n$ feature, and yielding a complex with $+det$,~~$-case$~~ at its root. #en (space en-dash)
Note that we could have combined *the* with *guy* as well. For the next Merge operation, one might combine either *the* with *guy* or *drank* with *the wine*, selecting the $+det$ feature and canceling the $-case$ requirement corresponding to the noun phrase argument *wine* – this corresponds to a conventional verb phrase. The root of this complex still has two unmet feature requirements: it selects a noun ($=n$), and assigns a case feature ($+case$). Note that an attempted Merge of *drank* with *the* before a Merger with *wine* would be premature: the v, $=det$, features would be percolated up, to a new *wine-the* complex. Now *wine* could no longer be combined with *the*. (On the other hand, there is nothing to syntactically block the sentence form, *the wine drank the guy*; presumably, this anomaly would be detected by the conceptual–intentional interface.)

Proceeding then with this path, depending on whether *the* and *guy* had been previously merged, we would either carry out this Merge, or, for the fourth and last step, Merge *the guy* with the verb phrase, in the process canceling the~~$-case$~~ feature associated with *the guy*. At this point, all formal features #en (space en-dash)
have been eliminated, save for the v feature heading the root of the sentence, corresponding to *drank* (in actual practice this would be further Merged with a tense/infl(ection) category). We can summarize the generation process as follows:

1. Merge 1: combine *the* and *wine*, yielding *the wine*.
2. Merge 2: combine *drank* and *the wine*, yielding *drank the wine*.
3. Merge 3: combine *the* and *guy* yielding *the guy*.
4. Merge 4: combine *drank the wine* and *the guy* yielding *the guy drank wine*.

Summarizing, Merge works on the model of chemical valency and feature cancellation. The core idea is that Merge takes place only in order to check features between its two inputs—a functor that requires some feature to be discharged, and an argument that can receive this discharged feature. The feature is then eliminated from further syntactic manipulation. After any Merge step, if a feature has not been canceled by a functor–argument combination,

that feature is copied to the root of the combination and further Merges attempted until we are left with only phonological and logical form features. After exhausting all possible Merge sequences, if any nonphonological or LF features remain then the derivation is ill-formed.

4.3.2.2 *Minimalism and movement* So far we have described only how a simple sentence is derived. Following Kitihara, as described in Epstein (1995), one can see that displacement or movement can be handled the same way, as a subcase of Merge. Figure 4.4 shows how. Suppose one forms the question, *What did the guy drink* by moving *what* from its canonical object position after the verb *drank*. Recall that we may define Merge as Merge(X, Y), where X and Y are either words or phrases. If X is a hierarchical subset of Y (roughly, a subtree), then this is a case of movement, as illustrated in the figure: $X =$ *what* is a subtree of $Y =$ *the guy drink what*. As usual, Merge forms a new hierarchical object, selecting and projecting one of the items, in this case *what*, as the root of the new tree. As usual, we must assume that Merge is driven by feature checking: we assume that there is some feature, call it Q for "question", that attracts *what*, while *what* possesses a $-Q$ feature as before, *what* moves to the *closest* position where its feature may be checked. Note that movement now amounts to copying the displaced element to its new position, forming literally *what the guy drink what*. Presumably a general phonological principle at PF avoids pronouncing *what* a second time, yielding the sentence that actually surfaces beyond the PF interface.

As we shall see in the next section, this approach also accounts for several of the formerly stipulated properties of movement. Perhaps more surprisingly, the notion of merge-as-derivation suffices to fix precisely the syntactic

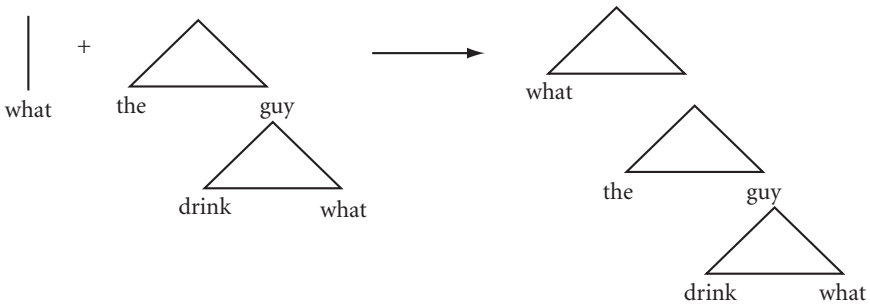

FIGURE 4.4 Movement or phrase displacement as a subcase of Merge. In this figure, *wh*-question formation is depicted as the Merger of two elements, *what* and a (traditionally named) sentence phrase, *the guy drink what*. Details about the inflection of *drink* and insertion of *do* to bear inflection information are omitted

relations appearing in natural languages, in this sense deriving a complex phenotype from a much simpler genotype.

4.3.3 *Deriving syntactic relations and constraints from Merge*

As described in the introduction, natural languages are characterized by certain specific properties, syntactic relations obtaining only among certain syntactic elements, under certain circumstances. These are seemingly forced by the minimalist framework and Merge itself. Let us review these properties here, showing how each follows from Merge without further stipulation. This of course is the key to establishing that we do not need separate, adaptive accounts for each of these properties, only for Merge.

- *Recursive generative capacity* This is a basic inherent property of Merge. Since Merge can apply recursively to its own output, indefinitely large hierarchical structures can be generated.
- *Structure dependence* Algebraically, Merge works via the concatenation of two (structured) objects. It is therefore a noncounting function: its inputs can be any two adjacent elements, but by definition it cannot locate the first auxiliary verb *inside* a string of elements (unless that element happens to appear at the left or right edge of a phrase), nor, *a fortiori*, can it locate the third or seventeenth item in a string. Note that given a "conceptually minimal" concatenative apparatus, this is what we should expect: clearly, Merge could not operate on a single argument, so the minimal meaningful input to Merge is two syntactic objects, not one or three.
- *Binary branching phrases* Since Merge always pastes together exactly two elements, it automatically constructs binary branching phrase structure.
- *Displacement* Given Merge, the previous section showed that a mechanism to implement displacement exists. Again, whether and how a particular human language chooses to use displacement is an option dependent on the features of particular words (up to the constraints enforced by Merge). For example, English uses displacement to form *wh-* questions, given a *Q* attractor in C(omplementizer) or root position, but Japanese does not. If displacement is a subcase of Merge, then the following constraints on displacement follow—constraints that are all in fact attested.

  - Displaced items *c-command* their original locations. C-command is the basic syntactic notion of scope in natural language; for our purposes, c-command may be defined as follows (Reinhart 1978):
    A c-commands B if and only if

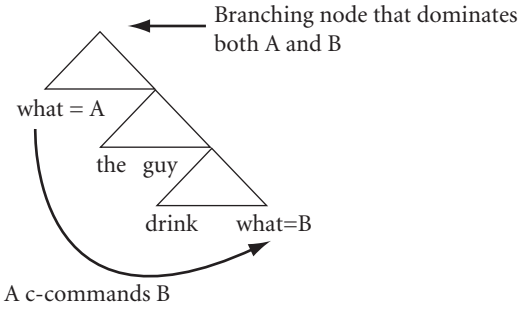

FIGURE 4.5 C-command is an asymmetrical grammatical relation between two hierarchical nodes in a sentence structure. This example shows that displaced elements always c-command their original locations

1. The first branching node dominating A dominates B
2. A does not dominate B
3. A does not equal B

Figure 4.5 illustrates. As one can see, in our displaced question sentence the first *what* (= A) c-commands the second *what* (= B), the object of the verb, because the first branching node above *what* dominates (lies above) the second *what*. Note that the c-command relation is asymmetric: the second *what* does not c-command the first.

The c-command relation between displaced elements and their original locations follows from a general property of Merge: given any two inputs to Merge, X and Y, where X selects Y, then X c-commands Y and all the subcomponents of Y, because by the definition of Merge we always form a new hierarchical structure with a label dominating *both* X and Y. In particular, for displacement, when X is a subset (conventionally, a subtree) of Y, the displaced X must dominate the original location that is a subpart of Y. Below, we show how to derive the form that c-command takes from more primitive properties of Merge.

- *Locality conditions* Displacement is not totally free, because feature checking is local. What blocks question-formation such as *What do you know how the guy drank*?, while allowing *How do you know what the guy drank*? This too has a direct answer, given Merge. Note that any phrase such as *How the guy drank what* is "locally convergent" in the sense that all its case and other feature-checking requirements have already been satisfied—this is what is called in linguistics an adjunct phrase. In other words, *How* satisfies any feature-checking requirement for the full sentence's aspect.

Another way to think of the same situation is that at this particular point in the derivation only phonological and semantic features remain in this subphrase. Therefore, this phrase may already be shipped off to LF—spelled-out—and is thereby rendered opaque to further syntactic manipulation. If this is so, then there is nothing that allows *what* to participate in further Merges—that is, it can no longer be displaced or moved. In contrast, the hierarchical object corresponding to *did the guy drink what* is still open to syntactic manipulation, because (in English) the aspectual/question feature associated with the full sentence has yet to be satisfied—and in fact, can be satisfied by Merging *what* with the sentence, moving *what* to the front: *what did the guy drink what*. Finally, such a sentence may be combined as an argument with *How do you know* to yield *How do you know what the guy drank*. In other words, given the local feature-checking driven properties of Merge, plus its operation on simply adjacent syntactic domains, we would expect locality roadblocks like the one illustrated.

To conclude our summary of how basic syntactic properties and relations can be derivable from the fundamental generative operator, following Epstein (1995) we can demonstrate that natural languages can express only a limited set of relations like subject-of, object-of, and c-command.

For example, the c-command relation holds between the subject noun phrase *the guy* and the object *the wine*, but not vice versa. Why? In so-called representational theories of syntax, such as government-and-binding theory, the notion of c-command is given by definition (Reinhart 1978). Its exact formulation is stipulated. However, c-command is derivable from properties of Merge and the derivational formulation presented earlier, as are the other basic syntactic relations.

To see why, consider again the Merge operation. Merge takes a pair of syntactic objects items and concatenates them. Syntactic structure is thus a temporal sequence of Merges, a *derivational history.* Given a derivational history and the sequence of syntactic structure the history traces out, we obtain the set of syntactically possible relations among syntactic objects. Let us see how. The derivation of our *wine* example is repeated below:

1. Merge 1: combine *the* and *wine*, yielding *the wine*.
2. Merge 2: combine *drank* and *the wine*, yielding *drank the wine.*
3. Merge 3: combine *the* and *guy* yielding *the guy*.
4. Merge 4: combine *drank the wine* and *the guy* yielding *the guy drank the wine*.

978–0–19–955327–3 04-Sciullo-c04 Di-Sciullo (Typeset by SPi, Chennai) 93 of 99 May 10, 2010 13:46
OUP UNCORRECTED PROOF – FIRST PROOF, 10/5/2010, SPi

Now the notion of a possible syntactic object and relation can be expressed via the following definitions.

Definition 1
Let A be a *syntactic object* if and only if it is a selected word or a syntactic object formed by Merge.

Definition 2
A syntactic object is said to *enter in the derivation* if and only if it is paired with another object via Merge.

Definition 3
We say A and B are *connected* if they are parts of another (larger, common) syntactic object C.

We can now *deduce* c-command from Merge:

Theorem 1
Let A and B be syntactic objects. A *c-commands* B if A is connected to B at the step when A enters into the derivation.

Proof sketch. Without loss of generality, let us see how this works with our example sentence. When *the* and *wine* are merged, they both enter into the derivation, and thus either may c-command the other, as is required. Merge creates a new hierarchical object, essentially the projection of *the*. Analogously, the verb *drank* and the object (the traditional object noun phrase) *the wine* c-command each other, because *drank* is connected to *the wine* at the time of their merger. These are the straightforward cases. The property that is more difficult to see is how one can derive the asymmetry of c-command. For instance, *drank* also c-commands all the subparts of *the wine*, namely, *the* and *wine*, but *the* and *wine* do not c-command *drank*. This is because at the Merger step when *drank* entered the derivation it was *connected* to *the* and *wine*. But the converse is not true. At the time when*the* and *wine* entered into the derivation (when they were Merged to form *the wine*), *drank* was not yet part of the derivation, hence was not visible. Hence, *the* and *wine* do *not* c-command *drank*, as is required. Similarly, the subject phrase *the guy* c-commands *drank the wine* and vice versa—because these two objects are Merged. Letting A = *the guy* and B = *drank the wine*, we see that the subject noun phrase is by definition connected to all the subparts of *drank the wine* because it is connected to them at the time it enters the derivation. Therefore, the subject c-commands these subparts, as required. The converse is not true—neither drank, nor the, nor wine c-commands the subject—because for A = *wine* for instance, A was not connected to the subject *at the time* it entered into the derivation. #

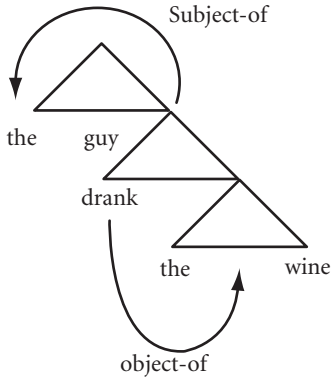

FIGURE 4.6 The core syntactic relations are fixed by visibility at Merger time. This figure depicts the *subject-of* and *object-of* relations, with the 'selected' or functor-like pair of the Merger drawn as the source of the arrow.

Indeed, it appears that if we take our definitions as specifying syntactic visibility, then all other syntactic relations reduce to subcases of the same criterion. Figure 4.6 illustrates the possibilities.

- *Object-of* is the relation: Merge and select a word base (functor) with either another word or a hierarchical structure.
- *Subject-of* is the relation: Merge a previously-merged hierarchical structure with a second hierarchical structure, selecting the first element as the new hierarchical root, and the second as the 'subject' (left-to-right order irrelevant—that is, the subject can appear either to the right or to the left).
- *Head-of* is the relation already described as the projection of features after Merge.
- No other (natural) syntactic relations are expected to be attested in natural languages, e.g., *subject-object-of*, relating, say, *guy* to *wine*, since these items are not *connected* at the time of their mutual participation in Merge.

4.4 From Merge to Language Use

A Merge-based model also meets a psychological-fidelity requirement for efficient language processing and accurate breakdown processing, beyond the broader kinds of language breakdown just described. There is a natural, transparent relation between a Merger sequence and the operation of the most general kind of deterministic, left-to-right language analyzer known in computer science, namely, the class of LR parsers or their relatives, as we demonstrate below. In other words, given that the general hierarchical Merge operator

978–0–19–955327–3 04-Sciullo-c04 Di-Sciullo (Typeset by SPi, Chennai) 95 of 99 May 10, 2010 13:46
OUP UNCORRECTED PROOF – FIRST PROOF, 10/5/2010, SPi

forms the basis for natural language syntax, then an efficient processor for language follows as a by-product, again without the need to add any new components. Of course, as is well known, this processor, like any processor for human language, has blind spots—it will fail in certain circumstances, such as garden path sentences like *the boy got fat melted*. However, we can show that these failings are also a by-product of the processor's design, hence indirectly a consequence of the Merge machinery itself. In any case, these failings do not seem to pose an insuperable barrier for communicative facility, but rather delimit an envelope of intrinsically difficult-to-process expressions that one then tends to avoid in spoken or written speech (Chomsky and Miller 1963).

First let us sketch the basic relationship between Merge and efficient LR parsing; see Berwick and Epstein (1995) and Stabler (1997) for details and variations on this theme. The basic insight is simple, and illustrated in Figure 4.7: a merge sequence like that in the figure mirrors in reverse the top–down expansion of each Rightmost hierarchical phrase into its subparts. Thus, since parsing is the inverse of top–down generation, it should be expected to follow nearly the same Merger sequence 1–4 as in Figure 4.7 itself, and it does. Consequently, all that is required in order to parse strictly left to right, working basically bottom-up and building the Leftmost complete subtree at a time, is to reconstruct almost exactly the Merger sequence that

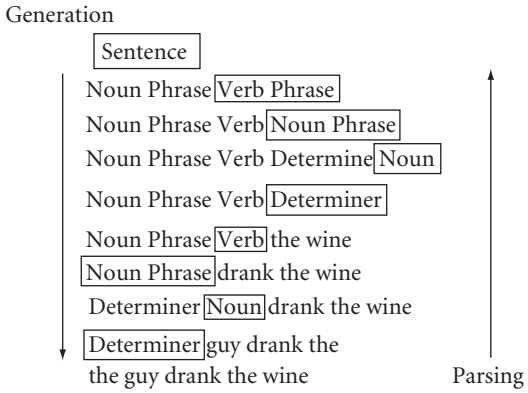

FIGURE 4.7 LR parsing is the mirror image of a top-down, right-most sentence derivation, and mirrors the Merge sequence for a sentence. This figure shows a line-by-line derivation for *the guy drank the wine*, where the boxed portion of each line shows that we expand the rightmost possible portion at each step in a top-down generation. Naturally, in a bottom-up parse, we reverse this process, and recover the leftmost complete hierarchical structure (the boxed portion) at each step.

generated the sentence in the first place. We assume in addition that if there is a *choice* of actions to take, then the processing system will again mirror the grammar, and so favor the economy condition that the closest adjacent feature should be checked, rather than delaying to a later point in the derivation.

Such a parser can work with a simple push-down stack, and has just two possible operations: either *shift* a word (a feature bundle) onto the stack, analyzing its features; or *reduce* (that is, Merge) the top two items on the stack, yielding a new hierarchical structure that replaces these items on the stack, forming what is traditionally known as a 'complete subtree.' Figure 4.8 shows

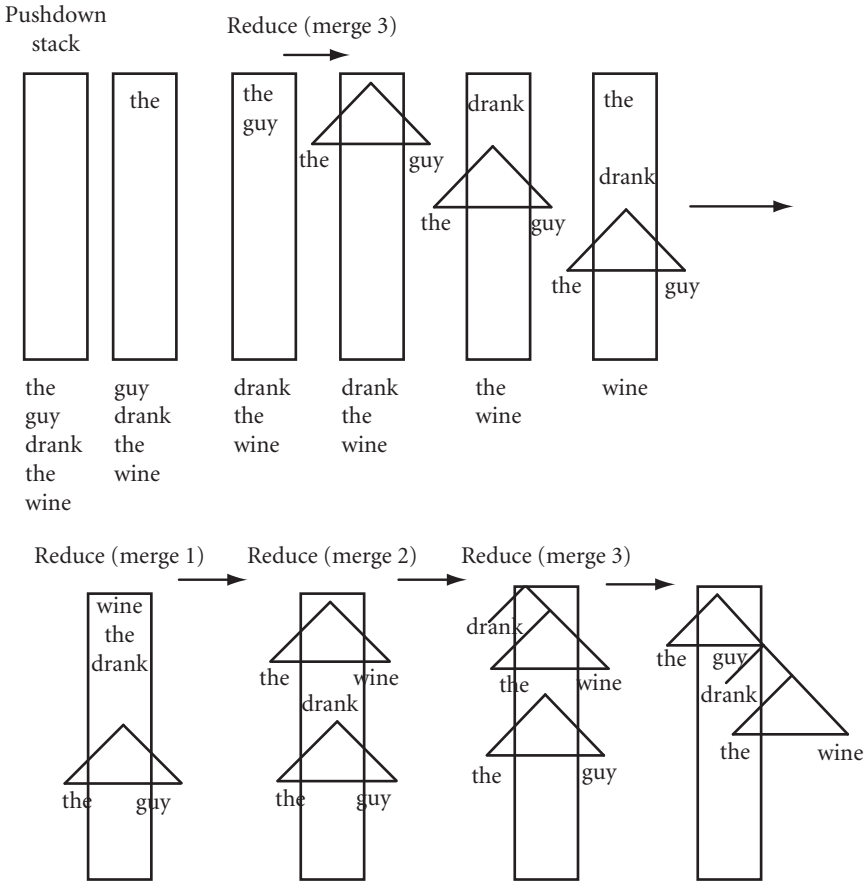

FIGURE 4.8 This figure shows how a step-by-step LR parse for the sentence *the guy drank the wine* mirrors Merger steps 1–4 for the same sentence. Each reduction corresponds to a Merge step.

the blow-by-blow action of such a machine operating on our example *drank* sentence.

As each complete subtree is produced, we envision that it is shipped off to the conceptual–intentional component for interpretation, up to the as-yet-unmet features still left at the top of each hierarchical structure. Recall that this is possible for locally convergent categories. For example, after *the* and *guy* are analyzed and then Merged, all the internal features of *the* and *guy* have now been accounted for, aside from those that play a role *external* to the entire phrase, such as the phrase's thematic role—but these are precisely any features that have not yet been "canceled" and are percolated to the head of the phrase for further processing. Thus these individual words may be interpreted. This proposal of incremental interpretation is essentially that found in Berwick and Weinberg (1986), and improved by Reinhart (2006). The basic principle amounts to what Chomsky (2001) called "derivation by phase:" the bottom–up construction of complete thematically satisfied syntactic units.

We can trace through the parse of our example sentence in detail as follows, relating Merge to parsing actions.

Step 1. Shift *the* onto the stack, recovering its features from the lexicon. (From now on, we shall omit the phrase *recovering its features from the lexicon* for each shift.)
Step 2. Shift *guy* onto the stack, on top of *the*.
Step 3. Merge 1: combine *the* and *guy*. Parser action: reduce *the* and *guy* to a complete phrase (leftmost complete subtree) replacing *the* and *guy* on top of the stack with any uncanceled, projected features.
Step 4. Shift *drank* onto the stack.
Step 5. Shift *the* onto the stack.
Step 6. Shift *wine* onto the stack.
Step 7. Merge 2: combine *the* and *wine* into a new hierarchical object, replacing both on the stack (this is the object of the sentence). Parser action: reduce.
Step 8. Merge 3: combine *drank* and the object into a new hierarchical structure, traditionally known as a verb phrase, *drank the wine.* Parser action: reduce.
Step 9. Merge 4: combine *the guy* and *drank the wine* into a complete sentence. Parser action: reduce. The parse is complete.

In many cases the choices for either shift or reduce (Merge) are deterministic, and allow such a device to work in the fastest possible time, namely, linearly in the length of the input sentence; but as is well known, in order to handle the ambiguity present in natural language, we must generalize an LR machine to work in parallel simply by carrying along multiple possibilities;

978–0–19–955327–3 04-Sciullo-c04 Di-Sciullo (Typeset by SPi, Chennai) 98 of 99 May 10, 2010 13:46
OUP UNCORRECTED PROOF – FIRST PROOF, 10/5/2010, SPi

there are known efficient algorithms for this (see Tomita 1986). In other cases, choices can be resolved by appeal to the local feature checking or economy condition imposed by the grammar; this leads directly to an account of some known language processing blind spots. Consider as one example the reflexive attachment of *yesterday* in sentences such as *John said that the cat will die yesterday*, described in the introduction. Why does the human sentence processor work this way? If in fact Merge proceeds by the most local feature cancellation at each step, then the answer is clear: *yesterday* can be merged with the lower verb *die*, so this choice is made rather than waiting for so-called late attachment—and this occurs before the *die* verb complex is shipped off for semantic interpretation. Hence, this is an operation that should be impervious to semantic effect, as indeed it seems to be. Similarly, such an approach also accounts for familiar cases of garden-path sentences, such as *the boy got fat melted*. Here too the basic situation, putting to one side many complexities, is that the noun-verb combination *boy got* is Merged "too soon" and taken as to be the main sentence—a processing error that we attribute to the local character of feature matching. It remains to see whether all psycholinguistic blind spots of this kind can be accommodated in the same way.

4.5 Conclusions

Taking stock, we see that Merge covers much ground that formerly had to be assumed in traditional transformational generative grammar. Many fundamental syntactic particulars are derivative: basic skeletal tree structure; movement rules; grammatical relations like object-of; locality constraints; even the cyclic character of grammatical rules—all these fall into place once the fundamental generative operation of Merge is up and running. These features are no less than the broad-brush outlines for most of human syntax—so nothing here has to be specifically selected for in a gradualist, pan-selectionist sense.

Of course, Merge will have little or nothing to say about the details of word features particular to each language—why English has a question word that sounds like *what*, or why such a word in English has features that force it to agree with an abstract question marker, while this is apparently not so in Japanese. Similarly, Chinese has no overt markings for verbal tense. The different words and associated features each language chooses ultimately lead to different possibilities for "chemical combinations," hence different "chemical compounds" or sentence construction types. But there is no need to invoke an array of distinct rules for each language, just as there is no need to invoke different laws of chemistry, once the basic principles are known.

As Chomsky (1995) has remarked, echoing the structuralists, while universal grammar has a long history, nobody has ever assumed there would be a universal morphology. Different languages will have different words with different features, and it is precisely here, where variation has been known all along, that languages would be expected to vary. In this sense, there is no possibility of an intermediate language between a non-combinatorial syntax and full natural language syntax—one either has Merge in all its generative glory, or one has effectively no combinatorial syntax at all, but rather whatever one sees in the case of agrammatic aphasics: alternative cognitive strategies for assigning thematic roles to word strings. Naturally, in such a system that gives pride-of-place to word features, one would expect deficits in feature recognition or processing, and that these could lead to great cognitive difficulties; many important details remain to be explored here. But if the account here is on the right track, while there can be individual words, in a sense there is only a single grammatical operation: Merge. Once Merge arose, the stage for natural language was set. There was no turning back.