The Development of Language

Acquisition, Change, and Evolution

David Lightfoot
University of Maryland

BLACKWELL
Publishers

4

Gradualism and Catastrophes

Grammars, then, are real, biological entities represented in individual mind/brains. They characterize a person's mature linguistic capacity, or a significant part of it. On the other hand, a language, like Odysseus, turns out to be a mythical, imaginary creature. It may be a convenient and useful fiction, like Odysseus and like the setting of the sun, but in reality it is derivative, the aggregate output of some set of grammars. We shall see that a language is not a coherent, definable entity.

So, when traditional historical linguists speak of a language changing, somebody with a biological view of grammars takes a reductionist stance and thinks of individual grammars changing and the changes spreading through a population. When others ask why a language should have changed in some way, we ask why grammars should have changed, aiming to explain the complex by the simple. This shift in perspective makes a big difference.

4.1 Grammars and Change

Let us consider how the biological view of grammars affects how we think about diachronic change, and let us distinguish it from other views of a grammar commonly invoked in discussions of language change. A grammar grows in the mind/brain of an individual upon exposure to relevant childhood experience. The grammar interacts with other aspects of a person's mental makeup, in a modular conception of mind. Different experiences may give rise to different grammars in different individuals, but it is a plausible initial assumption that grammars are subject to the same principles, parameters, and learning constraints, which are common to the species.

Looked at from this biological perspective, there is no single grammar of English – rather, various grammars which exist in the minds of English speakers. Grammars may vary somewhat, even among people whom we

call speakers of English. Speakers of English are not uniform, and grammars exist for people, not for languages. Let us distinguish terminology from reality here: proponents of the biological view of grammars occasionally speak of "the grammar of French" to refer to the grammars of French speakers in a generalized way, a kind of shorthand which abstracts away from individual variation. This usage may be misleading, but, as I have noted elsewhere, it is comparable to references to the French liver, the American brain, or the Irish wit; nobody believes that there is such an entity, but sometimes it is a convenient abstraction. Similarly, biologists standardly refer to "the fruit fly" or "the nematode," even in technical work, also abstracting away from variation which is irrelevant to the purpose at hand.

Grammars, then, are individual entities which exist in people and do not define languages as such. They exist in the minds of individual speakers, just as livers and brains exist in individual bodies. There is no such thing as "the grammar of English"; rather, there are thousands of speakers, all of whom have internalized grammars, some differing from others. That *set* of grammars generates much of the recorded body of what we call English, and much more that goes unrecorded.

Hermann Paul articulated this general view of language change clearly and fiercely in the nineteenth century. He emphasized the individual/biological view of grammars, noting in an early work that "dass die reelle Sprache nur im Individuum existiert" (actual language exists only in an individual) (1877: 325). His notion that "each individual has his or her own language" (1891: 23) is essentially the idea that there is no such thing as "the grammar of English." In attacking the group psychology of Lazarus and Steinthal (see chapter 2), he wrote in the original, 1880 German version of his famous *Principles* (1891): "Wir müssen eigentlich so viele Sprachen unterscheiden als es Individuen gibt" (we must in fact distinguish as many languages as there are individuals) (p. 31). His emphasis on the individual was a minority perspective and played no important role for a long period.

We can learn about the general principles, parameters, and learning constraints common to the species by studying how different grammars (in this biological sense) arise in children when the relevant experiences change minimally, from one generation to another. No two children have exactly the same primary linguistic data; they hear different things. Nonetheless, despite variation in experience, children often attain the same mature structural system. Initial experiences may vary indefinitely, but grammars show structural stability and vary only in limited ways. This structural stability is what enables us to communicate with each other. For Paul, variation and individuality were no mystery: "the problem which challenges solution is this: how comes it that while the language of each individual has its own special history, this degree of agreement . . . maintains itself within this miscellaneously constituted group of individuals?" (1891: 23).

Sometimes, however, minor changes in the relevant childhood experience cross a threshold and have consequences for the grammars which emerge. In that case, we have a different grammar, and this may entail far-reaching surface effects in terms of the class of sentences generated by the new grammar. The conditions under which grammars change can illuminate the general principles, parameters, and learning constraints, as we shall explore in the next few chapters. We investigate the change in grammars and ask how the change might have come about – that is, how childhood experience must have been different and what principles, parameters, and learning constraints were involved.

4.2 Social Grammars

Other linguists use the term *grammar* to refer to other, very different objects: pedagogical tools, inventories of morphological properties of languages, and, in a common usage, descriptions of socially defined entities. On this latter, social view of grammars, people speak of "the grammar of French" or of some dialect of French to allude to an algorithm which supposedly characterizes some socially defined idea of French encompassing the language of Michel Platini, François Truffaut, Jacques Chirac, and many others. This view of grammars remains common among certain historical linguists. Grammars are taken to be devices which generate a corpus of sentences belonging to some socially or politically defined entity, not the knowledge of an individual person. Such algorithms, if they exist, would define the conventional idea of a language.[1]

Biological and social views of grammars are very different from each other. The distinction between an individual's grammar and the group notion of a language is analogous to the biologist's distinction between individual organisms and species, and to the distinction made by historians and social scientists between individuals and societies or social classes. Species and societies are group notions. There is nothing right or wrong about working with one or other notion, but there is a difference – a vast difference. Certain questions need to be dealt with at the level of an individual organism, and for certain purposes we can abstract away from individual variation and operate at a macroscopic level of aggregate species, societies, and languages.

Questions arise which only make sense in one or other view of grammars. On the biological view, countless questions arise about the internal properties of these grammars: how they are represented in brains, how they emerge in young children, and so on. None of these questions is relevant if one thinks of a grammar as a social entity. And some questions get construed very differently.

Take the question: Do adults change grammars? For proponents of a social definition of grammars, the answer is yes, and that is a simple certainty. Since a grammar characterizes the corpus of English expressions, adult innovations will change that grammar insofar as they change the corpus. However, on the biological view of grammar that I have adopted, it is a logical possibility but by no means self-evident. It is undisputed that individual speech often changes through the course of a lifetime – we can track this nowadays through letters and tape recordings. In particular, idioms and slang expressions are known to change very rapidly. But the fact that an individual's speech changes does not entail that the individual's *grammar* changes. There is more, much more, to an adult's use of language than his or her grammar. Certain forms may be produced more or less frequently over the course of an individual's lifetime, sometimes deliberately, sometimes subconsciously. Some forms may be incorporated into a person's usage without being incorporated into the productive system characterized by the grammar; they may be specifically learned as forms to be used in particular contexts. It is quite possible that adults change their grammars in the sense in which I am using the term, but I know of no argument to that effect.

Adult innovations constitute grammatical changes for social grammarians, but that does not mean that they also reflect grammatical change for biological grammarians, because the latter are working with a different conception of grammar. Since grammars have structural stability, they may or may not be changed in adulthood. Although adult innovations may not affect grammars, they are nonetheless significant for the biological grammarian: they reflect changes to the primary linguistic data (PLD), the input experience for the next generation of language learners. Adult innovations, then, constitute one reason why an individual might be exposed to PLD which differ from what his mother was exposed to. There are other reasons, of course, relating to population movement, language contact, and the constant, ever-demanding need for strikingly expressive forms.

If one believes that there is a single grammar which generates the corpus of sentences found in English or some variety of English, then questions about how that entity changes will have little to do with language acquisition by children. Changes will be explained not by considerations of individual psychology but by historical principles which are independent of individual minds; in chapter 8 we shall discuss some actual explanations of this type.[2]

Different views of what a grammar is give rise to different questions, but a stronger point needs to be made: it is not clear that there can actually be such a thing as a social grammar. The sentences of, say, Old English do not constitute a recursively enumerable set. That is, there is no single, finite algorithm which can be acquired by a child and which generates the sentences of Old English and no nonsentences (we shall see this when we discuss optionality in grammars, later in this chapter). There is no reason to believe that such an algorithm exists in any real sense: it cannot be a psychological entity represented in an individual brain. Nor is it clear whose sentence patterns or pronunciation it would characterize – not any particular individual's, evidently. There is no comprehensive, central archive of English or French available to be consulted by William Safire or the Académie Française in order to determine whether something is a sentence of English or of French; nor is there any "social matrix" in which grammars might be located, except in postmodernist fantasies. It is hard to imagine what other kind of existence it might have, beyond being an intellectual construct in a researcher's notebook. In that case, the *social* grammar of Old English, whatever it is, would not be a generative algorithm. Although some historical linguists assume some sort of informal social grammar, the assumption is usually tacit, and the grammar is not explicated.

I know of only one context in which social grammars have been explicitly argued for, and that is in the studies inspired by William Labov's work (1994) on "variable rules." However, Labov works with a psychological grammar which has a different domain of application; it is not a social grammar in the sense outlined above. His claim is that speakers' grammars are psychological/biological entities existing in the minds of individual speakers in pretty much the same sense that I have argued here, but that these grammars incorporate rules which contain variables.

Labov's variables are socially induced in the sense that if a speaker uses and/or recognizes two past tense forms, say *dived* and *dove*, *sneaked* and *snuck*, both forms are incorporated into the speaker's grammar, along with an indication of the probability of each particular form occurring. All this information, including information about different dialects, is learned on exposure to the alternating forms. And likewise for syntactic operations such as movement from one structural position to another. So the individual's grammar incorporates knowledge of varying forms and their probability of occurrence. What is at issue here is not whether the grammar is biological

or social; Labov's grammars are clearly biological. The difference between his grammars and the ones I have described is that his incorporate much more information. Labov builds into his model of a person's linguistic knowledge information about social variability. Later in this chapter we shall discuss an alternative and, I think, superior approach to variability in terms of coexisting grammars.

Even if it is productive to work with a nonbiological, social notion of grammars, believing that grammars hold of social groups of one kind or another, that view leads to very different questions about historical change and very different answers. For the moment, I shall treat social grammars as fictions; but, in any case, they would entail a very different research program.

Adopting a biological notion of grammars enables us to approach an old debate from a new angle. The debate concerns the question of whether linguistic change is gradual or catastrophic. This debate is very similar to an analogous debate among evolutionary biologists. Some biologists believe that species evolve gradually, others that they evolve abruptly. Biologists, of course, are dealing with genotypical change, whereas we are dealing with change in phenotypes, change in grammars. We shall follow the first woman writer to be elected to the Académie Française. In her novel *Memoirs of Hadrian*, Marguerite Yourcenar wrote:

> When . . . two ideas are in contradiction, be ready to reconcile them rather than cancel one by the other; regard them as two different facets, or two successive stages, of the same reality, a reality convincingly human just because it is complex.

We shall see that "language" change is indeed gradual in certain respects, but that grammars undergo sudden, abrupt change from time to time. The debate about the gradualness of change turns on the categories used, languages or grammars.

4.3 Gradualism, Imagined and Real

Everyday common sense rests to some degree on the notion of continuity and gradual change. The basic idea is that if there is a little distortion, the patterns, processes, and structures of life don't change very much. Similarly with language. The overwhelmingly most common view among historians is that language change is gradual. Languages change in piecemeal fashion,

and there are thought to be no major discontinuities, except, of course, where the textual record has significant gaps. Recently, some writers have been aggressive in denying that abrupt, catastrophic change *ever* happens, with many phenomena changing simultaneously (see Carden and Stewart 1988; Harris, forthcoming; Harris and Campbell 1995 (48ff, 77ff, etc.), Hopper and Traugott 1993).

Gradualism has pretty much had complete hegemony. However, I'll argue here that whether or not change is generally gradual depends on what units of analysis are employed and which lens is used. It is languages which change gradually; grammars are a different matter. If we use biological grammars as our unit of analysis, then abrupt change happens. In fact, it must happen. Some changes may be small-scale and therefore may appear to be gradual, but appearances can be deceptive.

If we think macroscopically, in terms of changes in sets of more or less unanalyzed phenomena, using a wide-angle lens, then change always seems to be gradual. My favorite illustration is a study reported by Fries (1940). English used to have object–verb order, and one said things like "John said that Susan apples ate" (with the direct object preceding the verb in the embedded clause); present-day speakers would say *John said that Susan ate apples*. Fries offered statistical data showing that Old English alternated between object–verb and verb–object order freely, and that "the order of . . . words . . . has no bearing whatever upon the grammatical relationships involved" (p. 199). He found that object–verb order occurred 53 percent of the time around the year 1000, and that it was "gradually" replaced by verb–object order, reducing to 2 percent by the year 1500. He provided one set of statistics for each century but offered no analysis.

Fries's counts ignored the distinction between embedded and simple, unembedded clauses, and he had no analysis of the fact that the finite verb often appeared in second position in simple clauses. If we make such distinctions, we see that Old English grammars had object–verb order underlyingly and an operation moving finite verbs to yield verb-second order in main, unembedded clauses (van Kemenade 1987). Consequently, we find object–verb order uniformly in embedded clauses, but not often in simple clauses. Kroch and Taylor (1997) show that there was a dialect difference involving movement of finite verbs: in Middle English northern dialects moved verbs more than southern dialects.

In fact, two changes took place at the grammatical level: underlying object–verb order was replaced by verb–object order, and the verb-movement operation was lost. The *grammatical* change consisted in a change in the head-order parameter and the loss of northern "verb-second" grammars, each of which were abrupt. We'll see the details in chapter 6, but my point

here is that if our units of analysis are as gross as Fries's, change will look gradual.

At the other end of the scale is the microscopic approach, using a long, telescopic lens. The speech of no two people is entirely identical. It follows naturally that if we take manuscripts from two periods, even two periods in the lifetime of one individual, we shall be able to identify differences and so point to language "change." In this sense everything is in flux, and languages are constantly changing in piecemeal, gradual, chaotic, and relatively minor fashion – an example of the world of Heraclitus. Again we see gradual change. But the fact that a person's speech changes does not mean that his or her grammar changes.

Where historical records are rich, it is difficult to distinguish marked discontinuities merely by looking at the texts. In fact, nobody would have argued that linguistic change is abrupt and bumpy just by looking at textual records. Where the records seemed to suggest some bumpiness, it was assumed that fuller records would fill in the gaps and provide gradual curves. Sometimes those intermediate stages were reconstructed, as if there were an independent way of knowing what happened when the records were silent.

In this respect, the gradual/catastrophic debate among historical linguists differs from the analogous debate among evolutionary biologists. There gradualism has also been the dominant view, but the fossil record generally does not show gradualness. Rather, discontinuities are the norm, and punctuated equilibrium is overwhelmingly reflected in the fossil record. Consequently, proponents of abrupt change, punctuated equilibrium, derive very direct support from the relevant records (Eldredge and Gould 1972). Evolutionists committed to gradualism need to argue that the gaps in the fossil record are accidental. In language, the textual record, where rich and viewed with no analytical spectacles, has often suggested that speech changes gradually.

Not only do the texts suggest that change is gradual in the linguistic domain, but some changes really do progress gradually. Initial experiences are never entirely the same for any two speakers, and they may differ from each other in minor and insignificant ways. So in that sense, if one looks at them diachronically, they may change gradually. These changes may even be cumulative to some extent, while grammars remain structurally stable and invariant. Some construction type might become more frequent, perhaps as a result of taking on some expressive function. This would reflect a change in the way in which grammars are used, but not in the grammar itself. Such changes in frequency do not reflect a change in grammars, but they do constitute a change in the PLD for the next generation of speakers.

Not only may primary linguistic data change gradually, but the very nature of language acquisition ensures a certain kind of gradualness under usual circumstances. Normally the output of a person's grammar is a significant part of the linguistic environment that triggers the emergence of the grammar of that person's child. This works against major discontinuities in the class of expressions and their associated meanings and guarantees some stability from one generation to the next; generally one does not find a grammar yielding fairly uniform object–verb order being replaced abruptly by a grammar yielding uniform verb–object order. However, we shall see in chapter 6 that sometimes population shifts entail that parents' speech may be very different from that of their children, and one may find big discontinuities.

As historical linguists recorded and tried to understand the flux in the textual record, they overcommitted themselves to the proposition that change is necessarily gradual. This overcommitment takes the form of claiming that *grammars* themselves change only gradually. The internal properties of grammars, it is said, may differ only slightly from one generation to the next. A pair of rules may come to be reordered (Klima 1964; Harris 1980), or features may change in one generation (Allan 1988; Lieber 1982), or conditions on rules may be reformulated (Hausmann 1974). In a very early work (1969), Traugott formulated "diachronic processes," which mapped one grammar into another, and discussed the formal properties of those "processes," as if there was a formal mapping from one internal system to another. The idea here was that grammars, in general, may differ to a certain extent, but that grammars of adjacent generations of speakers may differ from each other in only minor ways. There were supposed to be formal constraints on the ways in which adjacent grammars may differ, and those formal constraints were to constitute a theory of change, quite independent of a theory of grammar.

This kind of gradualism has never amounted to much, in my view, but it is widely advocated. It is sometimes characterized through a theory of language types and sometimes through lexicalist models of grammar. In an approach which dominated work on syntactic change in the 1970s, the self-styled "typologists" distinguished "consistent" from "transitional" language types. Consistent languages exhibited harmonic properties, and those properties were mixed in various ways in transitional languages. They held that languages progressed from one "consistent" type to another, losing and acquiring harmonic word-order properties in a prescribed, long-term order. A consistent subject–object–verb language might gradually acquire all the harmonic properties of a consistent subject–verb–object language

in the appropriate order. The new verb–object order would emerge before the new noun–determiner or noun–adjective order, etc. The hierarchies were elaborate, sometimes involving 15 or 20 word-order properties; but no single language's history ever manifested more than one or two changes in these relationships. People postulated many unattested stages to account for instances where a fairly complete typological change was thought to have taken place, as between Latin and the modern Romance languages. Not unexpectedly, these linguists conducted the most detailed work on reconstructed histories, where problems of attestation did not arise. The commitment to gradualness in this model was more an article of faith than an empirical result based on investigation.

Alternatively, gradual changes have sometimes been modeled in "lexicalist" theories of grammar, in which particular grammars differ from each other not in terms of harmonic properties or of settings of abstract parameters, but only in terms of features of individual lexical items. The parameters of variation among grammars are defined by the available features. As a particular verb is recorded in a new syntactic environment, a feature in its lexical specification is changed accordingly (Allan 1988; Lieber 1982). The models invoked have allowed unrestricted systems of features, including features determining whether the lexical item undergoes a particular syntactic operation. Their virtue is that they permit accurate codifications of what occurs in historical texts at any period. Their drawback is that they are powerful enough to be compatible with just about any set of data. This approach to change implies that language acquisition is highly data-driven, that children match their input, which may vary pretty much without limit. However, these models, being too powerful, underestimate the bumpiness of variation and cannot represent a person's mature linguistic capacity, because they are demonstrably not attainable by children.

One way of seeing this is to consider cases of obsolescence, where a verb ceases to occur in a certain construction, and this is codified by a new lexical feature of that verb. The problem is that this new feature is sensitive to the *absence* of some form, and in general children do not acquire their grammars and set grammatical parameters on the basis of negative data (information that certain forms do not occur), as we saw in the last chapter. We'll see in a few pages that changes involving obsolescence of forms are particularly enlightening about the nature of parametric variation among available grammars.

Unconstrained lexicalist models have been popular with historians, because they are powerful enough to state even the narrowest of generalizations; but they cannot be models of a person's mature capacity, attained through normal childhood experience. They are powerful enough to be observationally accurate, systematizing what is recorded in the texts at some point by ever-changing lexical features, but the models characteristically offer no explanations for the changes being recorded. They reflect too narrow a concern for accuracy. Accuracy is not enough and may, in fact, be misleading. These models miss the fundamental property that grammars must be learnable. They also ignore the fact that texts from some period – say, the year 1354 – offer only a miniscule record of what could be generated by grammars of that period; consequently, there is no special merit in a system which generates exactly the set of sentences in that corpus.

The problem lies not with the features as such, or with the lexicalist models as such, but with their expressive power. A constrained lexicalist model would allow only a narrow range of features functioning in a narrow range of ways. For example, current work within the Minimalist Program (Chomsky 1995) allows functional heads to have certain features, which force the movement of categories. If they define only a narrow range of options, then they will not be compatible with just any data set, and they will allow us to capture the bumpiness of variation.

4.4 Catastrophes

As we have seen, the textual record, looked at casually, has suggested to many people that linguistic change progresses gradually. However, the Victorian economist Alfred Marshall once warned that "the most reckless and treacherous of all theorists is he who professes to let the facts and figures speak for themselves."

Considerations of language acquisition in children forced us in the last chapter to adopt an approach to grammars which predicts a bumpiness in variation and change. Natural languages are replete with partial generalizations which do not hold all the time and therefore cannot be learned inductively; one example is that *is* may be pronounced *z*. Consequently, linguists have postulated a rich linguistic genotype, UG, which provides the basic structures of grammars independently of any particular linguistic experience. Linguistic experience then refines the details and leads to the emergence of a mature grammar. That refining process consists in assigning words to a small number of categories, and that is one way in which grammars differ from each other, albeit within narrow limits. It also includes adopting abstract, structural properties. English grammars have a VP which includes a verb and then a noun phrase, $_{VP}$[V NP]; we say *ate an apple*.

Japanese grammars have the reverse order, noun phrase, then verb, $_{VP}$[NP V]; people say "an apple ate," pronounced *ringo tabeta*. Grammars like English are verb–object, and grammars like those of Japanese are object–verb.[3] Some grammars have an operation which raises verbs overtly to an independent inflectional position; others do not (section 6.3). Structural parameters like these introduce a kind of bumpiness into linguistic variation. Grammars which differ by a single parameter setting may generate very different classes of sentences, so speakers with these two grammars may be quite different from each other in terms of the class of sentences they use.

If grammars may vary in bumpy ways, then we would expect that they change in bumpy ways. That is, there must be points in history when grammars differed in terms of some structural parameter setting. In that case, changes are abrupt and catastrophic, with many surface effects. This follows simply from viewing language as manifesting a biological capacity represented in individual mind/brains, and then viewing change in the light of attempts to understand language acquisition as the emergence of these grammars in children under natural conditions.

Other changes in grammars may be small-scale (chapter 7). Lexical items may come to be categorized differently, and the recategorization may affect certain words before others, progressing in piecemeal fashion. In English, words like *must*, *can*, *may*, etc. were once more or less ordinary verbs, but they came to be categorized as instances of an inflectional element as a result of morphological changes which simplified the verb classes of Old English. Linguists have discussed this change intensively. The literary records suggest that the morphological changes affected some verbs before others. Likewise, the recategorization was piecemeal, with certain verbs being recategorized before others.[4] Each of the grammatical innovations, morphological and categorial, even if they are small-scale, must take place in direct or indirect response to changes in the PLD; there can be no other explanation.

Whether the changes are small-scale or large-scale, they constitute differences between grammars of different speakers. Those differences (which are what we mean by "change") do not have temporal properties, and therefore cannot be "gradual." Mark Hale put the issue well when he wrote to me that "the gradualness is a mirage created by our failure to distinguish between independent change events." He offers a good analogy. If we define *death* as the point at which some well-defined electrochemical activity in the brain ceases, it would make no sense to argue that death is gradual, either because it is followed by many further changes (e.g. changes in property ownership) which seem to be part of the same phenomenon, or because death moves from organ to organ gradually, or because it gives rise to "variation" in the population (dividing it into those who show the effects of death and those who do not). None of this would add anything to our understanding of death as a biological phenomenon.

The view I am advocating assumes that grammars hold not of languages, but of individuals. If there is no grammar of, say, Old English, then there can be no gradual shift in that grammar. In the same way, there can be no gradual shift in an aggregate French liver if alcohol ceases to be available, because that aggregate liver does not exist in any real sense. Grammars of individuals may differ from each other, just as individual livers may differ, perhaps as a result of differing alcohol consumption. If there are abstract, structural properties, then two grammars may differ in terms of one property, and that may entail many superficial differences; the set of sentences generated by those two grammars may differ quite significantly, although the grammars differ only in one property. This is true also for the grammars of two adjacent generations, where there has been grammatical change. In chapters 5–7, we shall see what sort of story this view of grammars allows for language acquisition and change.

I have found it helpful to think of change in grammars in the context of work on catastrophe theory. Catastrophe theory, developed originally by the French mathematician René Thom, is an attempt to provide a mathematical framework for modeling various kinds of discontinuous processes. For example, one can lower the temperature of a body of water, and a catastrophic change takes place at 0° Celsius, when it turns to ice. The water does not gradually become more ice-like: the phase transition is sudden. In chapter 2, I recommended John Casti's account of chaos theory; he also offers a good, balanced discussion of work on catastrophes (Casti 1994: ch. 2). He debunks the intuition that small, gradual changes in causes give rise to small, gradual changes in effects. The term *catastrophe* may be a little overblown to describe water changing to ice or the output of object–verb grammars changing to the output of verb–object systems; but Casti points out that the French *catastróphe* is not quite as catastrophic as the English *catástrophe* (p. 53). For us, "catastrophes" are the bumpy discrepancies that we find from time to time between the input that a child is exposed to and the output that the child's mature grammar generates. We shall see in chapter 6 that these discrepancies occur in a number of contexts, not just through the changes chronicled by historians of English, French, etc.

Casti illustrates elementary catastrophe theory (the only part that is in good shape mathematically), which deals with systems whose attractors are fixed points. One example is the gross national product (GNP), a measure

of the performance of the economy, which is determined by many factors – interest rates, money supply, consumer habits, industrial production, and so on. The level of GNP is a fixed-point attractor of the economic process.

> For every level of the inputs, the economy moves to a particular level of GNP, which can be envisioned as a point in the space of states of the economy. And since every setting of the inputs produces such a point, there is a whole surface of GNP points that the economy may produce – at least one for every level of interest rates, money supply, production facilities and all the rest. Catastrophe theory is designed to study the geometrical structure of this surface. (Casti 1994: 46)

Usually, if we change the inputs slightly, the corresponding level of GNP shifts only a little. However, sometimes we find combinations of input values such that a minor change entails a discontinuous shift to an entirely new region of the GNP surface. Such a value of the inputs is what is called a "catastrophe point."

> As it turns out, these catastrophe points arise at just those input levels where there is more than one possible fixed point to which the system can be attracted. And the jump discontinuity is a reflection of the system's "deciding" to move from the region of one attractor to that of another. Catastrophe theory shows us that there are only a small number of inequivalent ways in which these jumps can take place, and it provides a standard picture for each of the different geometries that the surface of attractors can display. (Ibid.)

Biophysicist Stuart Kauffman offers another example, which he takes to illustrate his notion of "the edge of chaos," a close cousin of the "self-organized criticality" findings of physicists Per Bak, Chao Tang, and Kurt Wiesenfeld.

> The central image here is of a sandpile on a table onto which sand is added at a constant slow rate. Eventually, the sand piles up and avalanches begin. What one finds are lots of small avalanches and few large ones. If the size of the avalanche is plotted on the familiar x-axis of a Cartesian coordinate system, and the number of avalanches at that size are plotted on the y-axis, a curve is obtained. The result is a relationship called a power law. The particular shape of this curve . . . has the stunning implication that the same-sized grain of sand can unleash small or large avalanches. Although we can say that in general there will be more tiny avalanches and only a few big landslides . . . there is no way to tell whether a particular one will be insignificant or catastrophic. (Kauffman 1995: 29)

Catastrophe theory deals with input values where the fixed point reflecting the system's behavior shifts from being a stable attractor to being an unstable one. This change in stability forces the system to move abruptly to a new stable fixed point. A small change in an input value leads to a discontinuous shift in the fixed-point attractor. This is not a predictive theory, and Casti shows that much of the bad press that catastrophe theory has received results from people interpreting it as a predictive system. It is explanatory, in much the way that evolutionary theory is explanatory, helping us to understand what has happened, rather than predicting what is to come.

The natural way for linguists to think of this is that different childhood experiences, different sets of primary linguistic data (PLD), sometimes cross thresholds, which entails that the system shifts, and that a new grammatical property results. So the inventory of variable properties constitutes the set of fixed-point attractors. This provides a productive way of understanding what happens in grammatical change and supports the viability of thinking of a small number of parameter settings as defining the structurally stable systems which we call grammars. We shall see in a moment that linguistic variation is typically not a matter of free variance, but rather oscillation between two fixed points of divergence.

The models of catastrophe theory strike me as particularly germane to work on grammatical change, and they may help us reach new levels of understanding. Conversely, work on abrupt change in grammars may help the mathematicians.

Why do mathematical theorems, obviously invented by human minds, apply so accurately to the outside world? This is the question posed by Eugene Wigner in his often cited paper "The unreasonable effectiveness of mathematics in the natural sciences." The answer lies in Paul Dirac's assertion that God is a mathematician. The cosmos has an astonishingly deep mathematical structure, which pervades many areas of nature. The same mathematics shows up all over the place. Juan Uriagereka (1998), working in the tradition of D'Arcy Thompson (1961), has been impressed by the prevalence of Fibonacci sequences in nature and thinks he has found them at work in grammars. He may be right, or he may be wrong; but it is exhilarating to find common principles at work in the structure of a flower's petals and in phonological inventories. These commonalities represent, in a sense, the deepest insights of science.[5]

4.5 Competing Grammars

Proponents of the biological view of grammars often write as if individuals have only one grammar. However, Tony Kroch and his associates (Kroch 1989; Kroch and Taylor 1997; Pintzuk 1991; Santorini 1992, 1993; Taylor 1990) have argued for coexisting grammars. They postulate that speakers may operate with more than one grammar, in a kind of "internalized diglossia." This work enriches grammatical analyses by seeking to describe the variability of individual texts and the spread of a grammatical change through a population.

In postulating two (or more) coexisting grammars in an individual, a researcher needs to show not only that the two grammars together account for a range of expressions used, but also that the two grammars are learnable under plausible assumptions about the child's trigger experience (PLD). In other words, diglossic grammars are subject to exactly the same learnability demands as any other biological grammar. There is no special issue here, and certainly no reason to believe that diglossic analyses are necessarily unlearnable, as is sometimes argued; nor is there any reason to reject the idea of diglossia because it might sometimes lead to postulating three or four coexisting grammars.[6] We cannot know in advance of investigation how many grammars speakers may have access to; that is precisely what we need to find out empirically. Learners have two or more grammars when their triggering experience leads to incompatible analyses. The difficulty here is not for the child but for the analyst: the analyst must decide when coexisting grammars must be invoked. However, we have no reason to believe that the world was designed in such a way as to make its investigation easy. Nor language. Kroch (1994: 184) notes that "when we reach the point where the linguist has as good a theoretical grasp on Universal Grammar as the language learner has an unconscious one, grammar competition will be as easily recognizable to the former as it already is to the latter."

In fact, this kind of diglossia represents an interesting approach to solving significant learnability problems. It offers a way of eliminating optionality in grammatical operations. Chomsky (1995) has argued that grammars do not permit optional operations. In that case, apparent optionality would be a function of coexisting grammars. Rather than allowing one grammar to generate forms *a* and *b* optionally, we would argue that a person has access to two grammars, one of which generates form *a*, the other form *b*; the speaker has the option at any given time of using one or other of the grammars. This move reduces the class of available grammars, eliminating those with optional operations.

Optionality within a grammar and optional access to two or more grammars may sound like the same idea in different clothing, but they are by no means notational variants. There is a big difference. Modern Dutch, German, Swedish, and other languages are verb-second, and necessarily so. I shall give an analysis of this phenomenon in chapter 6 (section 6.2); for the moment it is enough to say that the finite verb appears in second position, and that this is an obligatory requirement, not just one among a number of options. Dutch speakers say things like (1), not (2), where the verb follows two major constituents.

1 (a) $_{PP}$[In Amsterdam] $_{V}$live many good linguists.
(b) $_{NP}$[many good linguists] $_{V}$live in Amsterdam.
(c) $_{AP}$[Yesterday] $_{V}$visited we The Hague.

2 (a) *[In Amsterdam] [many good linguists] $_{V}$live well.
(b) *[Yesterday] [we] $_{V}$visited The Hague.

The obligatory nature of the phenomenon must reflect properties of UG, because it is not learnable; that is, it cannot be a function of the triggering experience of the Dutch or German or Swedish child. The reason is that if the child were free to develop an optional operation putting the verb in second position, it would take negative data to show the Dutch child that this was incorrect for Dutch and that the operation is obligatory. That is, the child would need information to the effect that having the finite verb in, say, third position leads to a sentence type which in fact does not occur in Dutch (2). Negative data of this kind are generally not available to children, as discussed in chapter 3; so the obligatory character of the verb-second property has to come for free, a function of the constraints of UG. So far we have a straightforward poverty-of-stimulus argument. The stimulus that children have, the triggering experience, is not rich enough to instruct the child that the verb-second property is obligatory; therefore we invoke UG.

Here comes the problem: Old English texts show verb-second order alternating with other orders. If we say that verb-second represents an optional operation in the grammars of Old English speakers, we lose our UG-based explanation for the obligatory character of the operation in the modern verb-second languages. We can avoid this problem by appealing to coexisting grammars: some Old English speakers had a grammar which generated consistent verb-second order, just like speakers of modern Dutch. Other speakers had a grammar in which there was no verb-second property,

and others had access to each of these distinct grammars. Thus there is no optional operation yielding verb-second structures in any grammar, and we therefore retain our explanation for the properties of the modern languages.

Similarly for the alternation between object–verb and verb–object order. In general, individual grammars do not manifest optional alternations of this type; further, they are precluded by UG for reasons similar to those which preclude an alternation between verb-second and verb-third order. Where a language has such an alternation, we say that this manifests diglossia, and that speakers have access to two grammars. Certain speakers have access only to one grammar; others have access only to the other grammar; and others have access to both grammars in an *internalized* diglossia. As long as each grammar is demonstrably learnable, then, far from raising problems of learnability, this kind of diglossia actually solves what would otherwise be a major learnability problem in the acquisition of optional operations. In general, grammars do not manifest optional, free alternations, and where languages have alternations, they are diachronically unstable and represent a transition whereby one of the grammars is driven into disuse.

Recall my earlier point that there may be no generative algorithm for socially/politically defined inventories. The case just discussed shows that in some cases a social grammar, were it to exist, would necessarily be incompatible with learnability demands, and therefore could not model an individual's capacity. If we write an algorithm generating these incompatible alternations, on the grounds that both forms occur as options in some single speech community, then we cannot write a learnable algorithm which captures the exclusiveness of the verb-second property in another community, like modern Dutch. As a result, there can be no single algorithm generating the options that we find in one speech community, if we are also to provide a learnable account of the exclusiveness that we find in another. For the purpose of accounting for acquisition and change, a "language" is not a good analytical tool; neither is a "grammar" which generates a socially/politically defined language.

On the view developed by Kroch and his associates, "change proceeds via competition between grammatically incompatible options which substitute for one another in usage" (Kroch 1994: 180). One reason for believing that this view of change through competing systems is along the right lines is that alternating forms cluster in their distribution, and the clustering follows from how sets of grammars unify the forms. We do not find free variance, but oscillation between two (or more) fixed points. This is reflected in the Constant Rate Effect of Kroch (1989).

Because parameters are abstract and structural, changing one parameter setting may entail a range of new surface phenomena. We shall see several examples of this in the next chapters. The Constant Rate Effect says that in such an instance all surface phenomena reflecting the parameter setting show usage frequencies changing at the same rate, but not necessarily at the same time. This is easy to understand if one grammar is replaced over time by another, and if that change takes place in a winner-take-all competition between the two grammars. Variation of this type does not stabilize and become a point of optionality within a grammar. We do not find complex arrays of linguistic data changing randomly. Instead, they tend to converge toward a relatively small number of patterns or attractors, as discussed in the context of catastrophes – in a kind of "antichaos" in the terms of Kauffman (1995). It is the theory of parameters which defines those attractors.

Kroch links the quantitative methods of sociolinguistics with the algebraic methods of theoretical syntax. His statistics would be quite mysterious if grammars involving abstract, structural parameters were not involved, restricting the variation to a narrow range. The common rate of change reflects the fact that abstract grammars are implicated. Multiple surface reflexes of single grammatical changes increase in frequency of use at the same rate, supporting the idea that the evolution of syntactic usage is controlled by changes in the underlying generative grammar.

For his simplest illustration, Kroch draws on work by Shawn Noble, who studied the replacement of the simple verb *have* by the complex *have got* in British English. Present-day speakers tend to use the complex form (3) where earlier speakers would have used the simple *have* (4).

3 (a) You've got brown eyes.
  (b) I've got a new job.

4 (a) She has what amounts to a high Cambridge degree.
  (b) They haven't the sense to come in out of the rain.

(3a) and (4a) indicate permanent possession, whereas (3b) and (4b) indicate temporarily bounded possession. The grammars, however, unify this distinction: one grammar has *have* meaning both permanent and temporary possession, and the other grammar has *have got* with both meanings. *Have got* replaces *have* across the board, regardless of whether it indicates permanent or temporary possession. The two meanings are distinct, and the temporarily bounded meaning is favored by approximately 0.65 to 0.35 in all three periods investigated by Noble. While the transition takes place,

Table 4.1 Effect of possession type on the choice between *have* and *have got*

Date	type	% *have got*	total	probability
1750–1849	temporary	12	83	.66
	permanent	4	108	.34
1850–1899	temporary	34	99	.64
	permanent	16	122	.36
1900–1935	temporary	89	74	.66
	permanent	70	43	.34

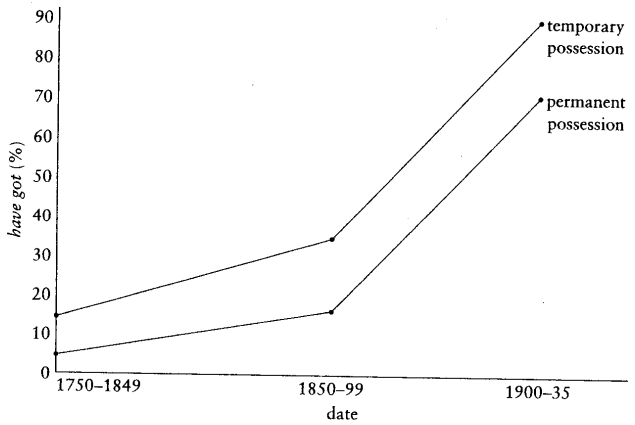

Figure 4.1 Graph for expression of possession type

many speakers use the new *have got* form for temporary possession and the old *have* form for permanent possession. Noble's figures are given in table 4.1.

The figures are shown in graph form in figure 4.1, where one sees that temporary possession occurs with the new *have got* form earlier than permanent possession, but the two types undergo the change at the same rate.

The distance between the two graphs indicates that at all times during the transition some individuals alternate between two grammars, using *have got* for temporary possession and *have* for permanent. This is presumably

due to factors independent of the grammars. One such factor might be the processing preferences which lead speakers to prefer one form to another in certain contexts when more than one option is available. An example of this involves the rise of periphrastic *do* in English, which we shall discuss in chapter 6 (section 6.3). Dieter Stein (1986) argued that during the period of their rise the forms with *do* are favored in contexts where its use eliminates certain complex clusters of consonants.[7] Mark Arnold (1995) has argued that *do* is favored where its use enables a transitive verb to stay adjacent to its complement; this would account for the use of *do* with transitive verbs – *Did Kim read the book?* – while intransitive verbs occurred without *do* – *Left Kim early this morning?* These preferences are not well understood, and, whatever they are, they have no effect on the overall progress of the change. They certainly do not entail that grammatically unified phenomena change independently of each other.

The fact that the two graphs proceed in parallel indicates that one abstract system in which temporary and permanent possession are treated alike replaces another; one grammar replaces another over a period of time. So the innovative forms are found at different frequencies, but they enter the language at the same rate. This is the Constant Rate Effect, and it shows up in several studies. It indicates that what changes over the course of time is the propensity of speakers to use one grammar as opposed to another in their language production. The unity of the change is a function of the abstract grammar.

If there is no optionality in individual grammars, then we may ask why and how one grammar replaces another in competition, rather than stabilizing into one system with an optional operation. The answer may lie in the Blocking Effect. Mark Aronoff (1976) argued that there is an economy restriction on lexical items such that morphological doublets do not generally exist; he called this the "Blocking Effect." From the point of view of acquisition, once a child knows that *went* is the past tense of the verb *go*, then it automatically follows from this economy restriction that *goed*, which most children experiment with to a greater or lesser degree, must not be a coexisting form, and it drops out of the child's usage. We do not generally add the productive suffix *-ness* to the adjectives *clear* and *bad* to yield **clearness* and **badness*, because there are preexisting forms *clarity* and *evil*, and the Blocking Effect precludes doublets of this type.[8] We need the economy restriction, and we need to say that it reflects a deep, UG principle, in order to explain the course of acquisition and the fact that doublets generally do not exist in the languages of the world.

However, while doublets generally do not exist, there are exceptions. We might take these exceptions too seriously and weaken the economy

principle to some kind of "tendency." However, this would be a mistake: if the no-doublets prohibition is not a principle, but only a tendency, it then loses its explanatory value. If it is just a tendency, then it needs to be explained and cannot itself be invoked as an explanatory notion. In the face of this, we might modify the principle so that it applies only where it holds absolutely. Or we might seek another principle which interacts with the Blocking Effect to yield the observed pattern. These moves are not promising, and Kroch (1994) has argued that the best approach is to retain the principle in its strongest form and to account for doublets, where they exist, sociolinguistically: doublets arise through language contact and compete until one form eventually wins out.

> Due to their sociolinguistic origins, the two forms often appear in different registers, styles, or social dialects; but they can only coexist stably in the speech community if they differentiate in meaning, thereby ceasing to be doublets. Speakers learn one or the other form in the course of basic language acquisition, but not both. Later in life, on exposure to a wider range of language, they may hear and come to recognize the competing form, which for them has the status of a foreign element. They may borrow this foreign form into their own speech and writing for its sociolinguistic value or even just because it is frequent in their language environment. Over time, however, as dialects and registers level out through prolonged contact, the doublets tend to disappear. (Kroch 1994: 185)

Initially, Kroch was obliged to make this move in order to retain the economy principle as an element of UG. This was needed in order to account for the course of language acquisition and for the relative rarity of doublets in the languages of the world. However, rather than being a last-ditch move to save the theory from embarrassing data, the idea of explaining theoretically inconvenient variation through differences in dialect and register turns out to have a new empirical payoff. Reconciling useful theoretical notions with initially contradictory data is, of course, a hallmark of successful sciences. In general, it is the most surprising reconciliations which are the most persuasive.

Kroch points to work on the history of English past tenses by his student Ann Taylor. There are many doublets of past tense forms, but they do not arise gradually, with equal frequency in all dialects at all times. There are occasional instances of new irregular, strong forms (i.e. where the vowel changes), like *dove* and *snuck* for the earlier *dived* and *sneaked*. However, the vast majority go in the other direction: new weak forms are introduced, exploiting the productive *-ed* ending, *walked* for *welk*, *awaked* for *awoke*, etc. Taylor (1994) showed that the appearance of doublets peaked

in the thirteenth to fifteenth centuries and was linked to the large-scale borrowing of northern features into written English, which until then was primarily a southern phenomenon. Northern dialects were affected by the large immigration of Scandinavian speakers into the north and north-east of England, sometimes in areas populated only sparsely by Saxons, during the various Danish and Norse invasions of the later Old English period. As adult Scandinavians learned English as a second language, so they simplified inflections – for example, generalizing *-ed* past tense forms. Naturally, this kind of second-language learning error would have occurred most often with less common verbs, so innovative weak forms are found mostly among the less frequent verbs.[9] Taylor found that doublets were much more frequent in the language of London (e.g. the writings of Geoffrey Chaucer), where there was a great mixture of dialects, than in the vernacular of the countryside (e.g. the letters of the Paston family, who were Norfolk gentry). All this would be mysterious if variation were not limited narrowly by grammatical theory.

When two communities encounter one another, members of one sometimes acquire the language of the other, learning it imperfectly as adults under less than ideal conditions. If one of the languages becomes dominant, then children may be influenced by changes introduced by the parents' imperfect learning. If that new form of the language is then passed on, a speech community may end up with more than one setting for a parameter. Children growing up in that community hear evidence for both settings and develop a diglossic capacity, using various forms to which they have been exposed, even though the forms may be grammatically incompatible. Over time, one form will win out, because of the economy demands of the Blocking Effect, which works against coexisting forms that are not functionally distinct. One example of this would be children in the Middle English period growing up in bilingual, English-Scandinavian communities.

There were competing past tense forms, and in some cases the innovative weak form has come down to present-day English, and in other cases it is the old strong form which has survived. In no case has a Middle English doublet survived to the present day, except in one context. In general, doublets do not last long. The average life span of doublets, measured by citations in the *Oxford English Dictionary*, is about 300 years, Kroch reports; but that figure is misleadingly high; it reflects citations in any dialect and takes no account of archaisms, where the form is used deliberately or unconsciously to replicate the usage of earlier times.

The one context in which doublets do persist is where the rival forms take on different functions. *Shined* and *fitted* persist in some dialects

alongside *shone* and *fit*, because they are causative while *shone* and *fit* are not (*He shined the light in her eyes* versus *The light shone in her eyes*; *She fitted him with a new suit* versus *That collar fit him last year*). Many people operate with two plural forms for *leaf*: we see *leaves* on a tree, but we watch the *Maple Leafs* at the hockey stadium in Toronto. The two forms are functionally distinct and refer to different things, leaves and a hockey team. Similarly, we may refer to our irritating colleagues as *Mickey Mouses*, not *Mickey Mice*, because, irritating as they may be, they are not mice and therefore may have a different plural form. If there are, say, three *Superman* films available, I might ask you how many *Supermans* you have seen, referring to the films and not to men, but I might refer to my department chair and my dean as *Supermen*. And my family are the Lightfoots, not the Lightfeet, because we are not feet. Here we find the limit of the Blocking Effect: it prohibits the coexistence of functionally equivalent items but allows them otherwise. That's why *badness* may exist alongside *evil* (see n. 8), because they mean different things.

The hallmarks of morphological doublets are that they are relatively rare. They tend to differ in register, in that only one of the forms is truly native, the other being a later accretion with a specific connotation. If they survive as a doublet, the two forms acquire distinct meanings, but one form is spontaneously overgeneralized in child language. This suggests that they manifest coexisting grammars in an internalized diglossia.

Kroch extends the approach to morphological doublets to account for language-internal syntactic variation, at least in circumstances where the properties of heads are involved. For example, when we find inflection-medial forms (5a) alternating with inflection-final forms (5b) in early English, then the two are considered functionally equivalent morphological variants, like the English past tense doublets. The idea here is that a finite verb like *visited* moved out of its VP to the inflection position (I), which might be medial or final (*e* is a "trace" and indicates an empty position from which a co-indexed item has moved, the verb *visited* in (5).

5 (a) ... that Kim $_{I}$[visited$_{i}$] $_{VP}$[e$_{i}$ London]
(b) ... that Kim $_{VP}$[e$_{i}$ London] $_{I}$[visited$_{i}$]

Like the past tense doublets, the inflection-medial and inflection-final forms compete with each other until one takes over entirely, yielding uniform word order.

The Blocking Effect is inviolable in language acquisition by young children, but it can be overriden as individuals learn a variety of styles and dialects. Over the course of time, however, the economy restriction on acquisition wins out over socially induced variation, unless the doublets acquire different meanings. This means that one grammar spreads through the population, and diglossia, including internalized diglossia, is eliminated. We now have an explanation for the unusualness and instability of apparent optionality, and the explanation is based crucially on the existence of competing, coexisting grammars.

4.6 The Spread of New Grammars

Assuming that individual speakers may operate with more than one grammar also permits a new approach to the replacement of one grammar by another across a society – that is, the spread of change through a speech community. Weinreich, Labov, and Herzog (1968) distinguish between innovation by the individual (which is typically abrupt) and spread across a community (which is typically gradual and manifested by systematic variation). Only the latter constitutes "change" in their view. Reserving "change" for spread of a phenomenon across a community makes little sense for the perspective focusing on biological grammars. However, we agree that individual change is abrupt. The spread of a change across a community raises very different questions for this perspective and is approached through the methods of population biology.

An account of language acquisition at the individual level leads naturally to an account of language change at the group (or population) level. An individual may be exposed to PLD which differ from the parents' PLD, because of population movement, language contact, adult innovations, or perhaps because the PLD happen to be truncated in some way – so do not include certain expressions or do not include them with the frequency of a generation earlier. Recall that the PLD are finite, consisting of certain robust, structurally simple expressions presented to a child, who sets relevant parameters. One individual may set some parameter differently from older people in her community; then it is likely that, because of the grammatical change, she will produce different utterances from other people in her community. These new expressions, in turn, affect the linguistic environment, and she will now be an agent of further change, by virtue of the fact that her younger siblings will have different PLD as a result of what she produces with her new grammar. As the younger siblings also set the relevant parameter in the manner of the older sister, so other people's PLD will differ. Thus a chain reaction is created. It is in this sense that grammatical change can spread analogously to what has been observed in population genetics, replicating aspects of evolutionary change.

Partha Niyogi and Bob Berwick (1995, 1997) have now produced a computer model which analyzes change in this way and *derives* the trajectory of changes. They postulate a learning theory with three sub-components: a theory of grammar, a learning algorithm by which a child generates grammars on the basis of exposure to data, and PLD. They postulate a population of child learners, a small number of whom fail to converge on preexisting grammars. After exposure to a finite amount of data, some children converge on the preexisting grammar, but others do not; they attain a different grammar.

> The next generation will therefore no longer be linguistically homogeneous. The third generation of children will hear sentences produced by the second – a different distribution – and they, in turn, will attain a different set of grammars. Over successive generations, the linguistic composition evolves as a dynamical system. (Niyogi and Berwick 1997: 2)

Language change, in this simulation, is a logical consequence of specific assumptions about the theory of grammar, the learning algorithm, and the PLD. Niyogi and Berwick produce a plausible model of population changes for the loss of null subjects in French. Interestingly, their model yields different trajectories for different changes. A common trajectory is the S-curve so familiar to historical linguists studying the spread of changes (e.g. Weinreich, et al. 1968, Kroch 1989). A change may begin quite gradually, then pick up momentum and proceed more rapidly, tailing off slowly before reaching completion.

The fact that changes progressing through populations can be graphically represented by a S-curve is not surprising to those who think in terms of chaotic systems and catastrophic reanalyses. The success of Niyogi and Berwick lies in showing that it is not impossibly difficult to compute (or simulate) grammatical dynamical systems. They show explicitly how to transform parameterized theories and memory-less learning algorithms into dynamical systems, producing results along the way. As a result, they *derive* the S-curve, rather than build it into their model as a specific assumption. Further, not all changes progress in S-curves; we find other trajectories – indeed, their model generates other trajectories for different changes. The model entails that changing specific elements of the theory of grammar or of the learning algorithm produces different trajectories for any given change. This means that their model may be amended in light of the way in which it matches the actual trajectory for specific changes in specific languages. This offers a new empirical demand for theories to meet, in addition to demands of learnability, coverage of data, etc.: theories should provide the most accurate diachronic trajectories for changes. There is more to be said about this work, and we shall return to it in chapter 8.

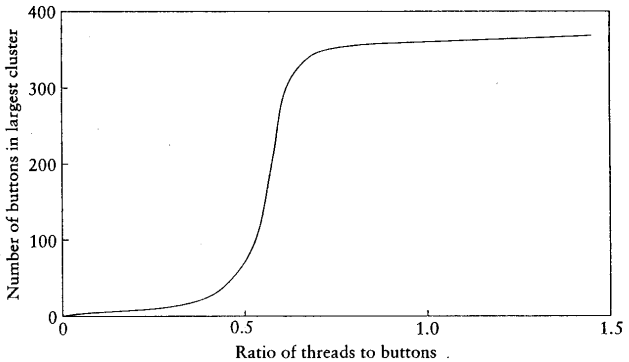

Figure 4.2 Phase transition. From Stuart Kauffman, *At Home in the Universe*, © 1995 by Stuart Kauffman. Used by permission of Oxford University Press, Inc.

This reminds me of a toy experiment recounted by Kauffman in the context of "random graphs." Imagine 400 buttons on a dining room table.

> Randomly choose two buttons and connect them with a thread. Now put this pair down and randomly choose two more buttons, pick them up, and connect them with a thread. As you continue to do this, at first you will almost certainly pick up buttons that you have not picked up before. After a while, however, you are more likely to pick at random a pair of buttons and find that you have already chosen one of the pair. So when you tie a thread between the two newly chosen buttons, you will find three buttons tied together . . .
>
> The important features of random graphs show very regular statistical behavior as one tunes the ratio of threads to buttons. In particular, a *phase transition* occurs when the ratio of threads to buttons passes 0.5. At that point, a "giant cluster" suddenly forms . . . As the ratio of threads to buttons passes the 0.5 mark, all of a sudden most of the clusters have become cross-connected into one giant structure. (Kauffman 1995: 56)

Kauffman shows the phase transition in the graph reproduced as figure 4.2.

Kauffman believes that this is a toy version of the phase transition which leads to separate water molecules freezing into a block of ice. The point of

using mathematical idealizations in science is that they help to capture the main issues. Kauffman's magical mathematics reveals the kind of pattern that we find in grammatical change and the kind of thing that Niyogi and Berwick have successfully modeled.

For the moment, we may note that Niyogi and Berwick have provided a model of how changes in individual grammars progress through a population of speakers. The fact that the model derives different trajectories for different changes, and that at least some of these trajectories seem to match the real, historical world reasonably well suggests a degree of plausibility. This is a remarkable result, which clearly could not be replicated under a social definition of grammars, which denies the usefulness of individual, biological grammars. The model should impress proponents of social grammars, who must postulate the S-curve as an unexplained primitive.

In discussing competing grammars, Kroch speaks repeatedly of changes progressing slowly and gradually over long periods. There may be slowness and gradualness at the level of the spread of the change through a population of speakers, but grammatical changes need to be instantaneous at the individual level. This follows from the Blocking Effect, which permits coexisting forms only if they are functionally distinct, as discussed earlier. We also know that some changes progress through a population rather rapidly. Earlier we noted that Kroch himself cites the fact that the average life span of doublets, as attested by the *OED*, is 300 years, but that this figure must be greatly exaggerated, since it ignores archaisms and geographical variation in dialects. This all strongly suggests that structural changes are rapid and abrupt at the individual level, and that, in many cases, they also spread through a population rapidly. The speed of the spread depends on many nongrammatical factors relating to social cohesion, facility of communication among different groups, and the like.

My own early work on grammatical change argued for abstract reanalysis on the basis of the dates of initiation and completion, as witnessed by initial and last attestations. For example, if a variety of phenomena drop out of the language at the same time, it might be plausible to claim that those phenomena all manifested a single grammatical change (Lightfoot 1979). Some of the early conclusions have turned out to be sound, whereas others have called for revision. In the early work, there was little concern with intermediate stages or with the spread of a change through a population. Now the work on competing grammars and on population models allows us to model transitions and enriches our understanding of change very considerably.

4.7 Parametric Change

In *As You Like It* (Act III, scene 2), Rosalind explains how time can amble, trot, gallop, or stand still for people in different circumstances. Historians have always known that languages sometimes undergo a period of rapid change, then settle into relative stasis. I have argued here that changes sometimes take place "catastrophically," and that grammars change abruptly. At certain points, dramatic changes take place, often simultaneously. It is natural to try to interpret a cascade of changes in terms of a new setting for some grammatical parameter, sometimes having a wide variety of effects. If we distinguish changes resulting from reanalyses, then we need to know what to look for in seeking grammatical changes. Catastrophic changes, resulting from a new parameter setting, have distinctive features and are quite different from the piecemeal, gradual, chaotic changes which constantly affect the linguistic environment. In earlier work, I identified six distinctive features of grammatical change (Lightfoot 1991: ch. 7).

First, each new grammatical property is manifested by a cluster of new phenomena. For example, the loss of the V-to-I operation in English (see chapter 6) entailed the predominance of forms like *Kim always reads the Bible* in place of the earlier *Kim reads always the Bible*, and the obsolescence of inverted and negative sentences like *Reads Kim the Bible?* and *Kim reads not the Bible.* These apparently unrelated changes took place in parallel, as demonstrated by the statistical studies of Kroch (1989), which showed the singularity of the change at the grammatical level (and led Kroch to postulate his Constant Rate Effect).

Second, not only is a new grammatical property typically manifested by a cluster of new phenomena; it also sometimes sets off a chain reaction. An example from the history of English is the establishment of verb–complement order. I showed (Lightfoot 1991) that this led indirectly to the introduction of an operation analyzing *speak to*, *spoken to*, etc. as complex verbs. Such chain reactions can be understood through the acquisition process: a child with the new verb–complement setting is induced by the constraints of UG to analyze some expressions differently from the way they were analyzed in earlier generations. As a result, the new grammar comes to differ from the old in more than one way.

Third, changes involving new grammatical properties tend to spread rapidly and manifest the S-curve discussed earlier. The old negative patterns associated with the V-to-I raising operation (*Kim reads not the Bible*) were robust and widely attested in the texts until their demise, which was rapid (see chapter 6). The fast spread of new grammatical properties is not surprising if one thinks of it in the context of language acquisition, as

I have just discussed. Once the linguistic environment has shifted in such a way as to trigger a new property in some children, the very fact that some people have a new grammatical property changes the linguistic environment yet further in the direction of setting the parameter in the new fashion. That is, the first people with the new parameter setting produce different linguistic forms, which in turn are part of the linguistic environment for younger people, thereby contributing to the spread of the new setting.

Fourth, obsolescence manifests new grammatical properties. When structures become obsolete, one cannot attribute their obsolescence to the ebb and flow of nongrammatical changes in the linguistic environment. A novel form may be introduced for expressive reasons, at first without affecting any grammar; but a form can hardly drop out of the language directly for expressive reasons or because of the influence of another language. On the contrary, obsolescence must be due to a structural domino effect, a by-product of something else, which was itself triggered by the kind of positive data generally available to children, as we saw in the context of the Blocking Effect.[10]

Fifth, any significant change in meaning is generally a by-product of a new grammatical property, for much the same reason that the obsolescence of a structure must be the indirect consequence of a more abstract change. In chapter 1 I discussed changes affecting the thematic roles associated with particular NP positions in the case of verbs such as *like*, *repent*, *ail* (the direct object of these verbs could once be an experiencer, while in modern English only the subject may be an experiencer; so people said things along the lines of "Apples like me" for the modern *I like apples*). It is hard to see how these changes could have arisen as idiosyncratic, nongrammatical innovations that somehow became fashionable within the speech community. Rather, such changes must be attributed to some aspect of a person's grammar, which was triggered by the usual kind of environmental factors – for these English "psych-verbs," the existence of only structural cases (Lightfoot 1991: ch. 6).

Sixth, new grammatical properties occur in response to shifts in simple data, with cues occurring in unembedded domains only; they are not sensitive to changes or continuities in embedded domains. Embedded domains are as likely as unembedded domains to reflect the usual to-ing and fro-ing of the chaotic linguistic environment, but they have no effect on the development of grammars in children. This follows from degree-0 learnability, the claim that grammars are learnable – that is, parameters are set – on the basis of data from unembedded binding domains (Lightfoot 1991, 1994).

All of this will make more sense when we discuss some actual cases of new parameter settings, in the next chapters.

So is change gradual or abrupt? I have argued here that it depends on the lens that one uses. Experience is generally fluid, and if one is concerned exclusively with describing differences in experience, in terms of what one finds in comparable texts of different generations, then change appears to be fluid, gradual, piecemeal, and chaotic. Similarly, if one thinks in terms of some kind of social grammar, where the "grammar" (mythical, I believe) is some kind of codification of the texts produced by some group of speakers, then again change appears to be gradual. Also, if one thinks in terms of a language changing, taking language to be a group phenomenon, change appears to be gradual, like water flowing. However, such commonsensical notions do not get us very far. For a plausible account of language acquisition, we must think in terms of abstract grammars; change is then a function of different grammars emerging in different people.

If we think in terms of biological grammars represented in individual mind/brains, then grammars differ from each other abruptly, a function of the limited variation allowed by UG. From this perspective, we see a series of snapshots – that is, distinct, individual grammars – and we expect languages to differ from each other in bumpy ways. This view may violate common sense, but physicists who view space in eleven dimensions know that common sense is not always a good guide. It provides a way of studying how grammatical shifts affect populations of speakers, an area where historical work has been bedeviled by confusion. One does not find one set of grammars being replaced by another set overnight; the textual record certainly does not suggest that English speakers had object–verb grammars replaced uniformly by verb–object grammars on some National Head-Change Day in the thirteenth century. Nor would we expect such a scenario if we recognize that speakers may operate with more than one grammar, and that grammar change may spread through a population over the course of time. Nonetheless, one does find that certain changes progress rapidly, and we can understand that phenomenon by postulating abrupt change at the level of individual grammars, and new grammars then spreading through a population of speakers.

The characterization of abrupt grammatical change sketched in this chapter makes sense only if we view a grammar as an individual, mental entity, and not as some kind of social entity codifying the data attested in the texts of some period. Making that distinction gives us a way of thinking about the spread of new grammars through a community using the methods of population genetics.

The success or failure of a research program is best determined by seeing whether it is productive or whether, like the neo-grammarian program of the nineteenth century, it leads to diminishing returns. In this chapter we

have looked at the speed of change fairly abstractly, and I have made several references to forthcoming arguments in chapter 5 and beyond. It is time to move on to chapter 5, to enter Polyphemus's cave, and to think more concretely. Let us now examine some specific instances of catastrophic change at the grammatical level and see how they enable us to understand the phenomena of language change.

Notes

1 Traugott and Smith (1993), among many others, conceive of a grammar in this way, thinking of it as a device to generate a corpus of sentences belonging to some socially or politically defined entity (Lightfoot 1995). In early work (1969), Traugott took quite an extreme position in this regard, seeking a "diachronic grammar," a single, formal device characterizing different forms and different historical stages of English. This is a very natural logical possibility – in fact almost inevitable – if one thinks in social terms; after all, why should the social unit be defined as just covering one point in time? For discussion, see Lightfoot 1979: 28–35.

Graffi (1995) discusses the nineteenth-century notion of social grammars which Paul inveighed against.

2 Several groups of researchers have attributed language change to general historical principles, not to changes in PLD. The "typological" explanations offered in the 1970s and the more recent studies on "grammaticalization" both invoke a general theory of change. Grammaticalization is a semantic process, whereby an item with a full lexical meaning comes to acquire a more abstract, functional, grammatical meaning. Those historicist explanations, of course, are not couched in a biological view of grammars, and analyses are formulated in very different terms.

However, some modern work which adopts the biological view of grammars also invokes general historical principles. This work builds into grammatical theory devices which incline grammars to change in certain directions. This work raises distinct and interesting issues related to the motivation for change, which I return to in chapter 8. I mention it here to show that there is no logical connection between the biological view of grammars and the contingent approach to change that I am adopting, which views structural change as due entirely to prior changes in the PLD rather than to general principles of change or to nineteenth-century-style laws of history.

3 [NP V] order may be base-generated, or it may result from some adjunction operation if grammars have V NP order uniformly in initial structure, as argued by Kayne (1994). I ignore the form of this parametric variation here.

4 Warner (1993) has argued that some changes relating to the English modals were first instantiated earlier than has been claimed. We shall discuss changes in modals in chapter 7.

5 This assumes a realist approach. The universe is patterned in ways that are described by mathematics; the mathematician discovers truths that are independent of his or her culture, and those truths are qualitatively different from traffic conventions or codes of etiquette. As Martin Gardner puts it,

> if all intelligent minds in the universe disappeared, the universe would still have a mathematical structure and in some sense even the theorems of pure mathematics would continue to be true. . . . For a mathematical realist a tree not only exists when nobody looks at it, but its branches have a "tree" pattern even when no graph theorist looks at them. (1996: 281)

This is very different from the postmodernist view that mathematics is a cultural construct, intelligible only within the context of culture. For the "conceptualists," mathematicians do not discover preexisting, timeless things like pi and dodecahedrons; they construct them. Gardner draws this distinction nicely and takes apart the conceptualist stance.

6 Traugott and Smith (1993: 437–8) have argued that diglossic grammars would be unlearnable, and Harris and Campbell (1995: 86) are alarmed at the prospect of "tri- or tetraglossia." There is no reason in principle why diglossic grammars should be unlearnable, although some particular proposals may raise questions of learnability. This just means that we have to be careful in our analyses. I discussed this point in the context of arguments by Susan Pintzuk (1991) for a diglossic treatment of Old English writers, arguing that her particular diglossic analysis is unlearnable (Lightfoot 1997b).

The notion of coexisting grammars is itself a high-level abstraction, but the work of Kroch and his associates is weakened by not discussing how competing grammars might be used by speakers. If their theory allows speakers to switch from one grammar to another in mid-VP, it is probably too powerful.

7 Stein studied the corpus of Shakespeare and argued that the periphrastic *do* is most common where the inflectional ending on the verb would yield a complex consonantal cluster. So *do* is most common in the second person singular, where the ending is *-st*; so *Thou dost sing* occurs more commonly than *I do sing* because it is an alternative to the complex *Thou sing'st*. The frequency of *do* is a function of the consonantal clusters to be avoided. Stein offers some figures: verb stems ending in a vowel occur with *do* 10 percent of the time; verbs ending in a *-k* 24 percent, and verbs ending in a *-t* or a *-d* 38 percent.

Do was well established by the time of Shakespeare, and it is unclear whether Shakespeare's usage reflects common speech at the time or whether it also reflects, to some extent, the poetic demands of euphony.

8 We do, however, use these forms in specialized senses, which are not equivalent to the preexisting *clarity* and *evil*. So linguists often talk of expressions which may not occur in some language as "bad." Then they can go on to compare the "badness" of *Who did you say that left Rio?* with the "badness" of *Who said you that left Rio?* These expressions may be bad in this sense, but they are not evil or immoral in any way. Also, *clearness* has a specialized meaning among Quakers.

9 This conforms to what we know about the storage of forms in the mental lexicon: irregular forms are stored as independent entries (this idea will be important in section 7.2), whereas regular forms are not stored independently, but rather as some kind of stem which then undergoes a productive operation to yield the occurring form. A nice piece of evidence for this division is given by Pinker (1994). In forming compounds, we access lexical entries. If an irregular plural, for example, has its own lexical entry, then we are not surprised to find compounds like *mice eater*, alongside *rat eater*; the compounding exploits the lexical entries *mice* and *rat*. However, we do not find **rats eater*, because *rats*, being a regular plural, is not a distinct lexical entry; therefore the form is not available to the compounding operation.

10 For a recent application of this methodology, see Warner 1995: 542.