4

GRAMMAR SIMPLIFICATION

In the preceding chapter we were primarily concerned with determining types of linguistic change in languages both living and dead. Four individual categories of change were isolated: rule addition, rule loss, rule reordering, and simplification. There do not seem to be other major kinds of change that cannot be reduced to one of these four, and indeed two of the four—rule loss and reordering—can be understood as belonging to simplification in its broadest sense. We shall find later that other types of apparently unrelated changes such as analogy often reduce to special cases of simplification.

It is possible that further investigation will turn up other types of change that do not belong in any natural way to any of the four primary types posited. If this happens, it will affect what has been said only in that we must ascertain its implications for our conception of change as change in the

speaker's competence—in the system of rules underlying his ability to speak his native language.

The presentation in Chapter 3 must not be wrongly construed. Although primary change is ultimately reducible to one of four types, one should not conclude that each of the four is equally probable given a particular grammar. Such does not seem to be the case. Rule addition seems characteristically to occur in the adult's grammar, whereas rule loss and reordering seem primarily to occur in the child's grammar. Simplification, rule loss, and rule reordering seem typically to occur in the transmission of language from generation to generation, not within the speaker's adult life span.

This chapter deals primarily with the processes of linguistic change between generations. We shall probe into the mysteries of change more deeply than in the preceding chapter, where we were merely moving about on the surface of linguistic change.

We shall develop here a unified picture of linguistic change as it is at present comprehended within the theory of language and grammar sketched in Chapter 2. Implicit in any theory construction is the setting up of hypotheses, and in this chapter we shall, for example, hypothesize that young children, and not adults, have the ability to construct an optimal (simplest) grammar from exposure to a finite set of speech performances. We shall hypothesize that the grammars of adult speakers change, if at all, by minor alterations relatively "late" in the set of ordered rules comprising a given component of the grammar. Adults, that is to say, are capable of incorporating innovations in their grammar, but not in general capable of redoing their grammar in ways open to children constructing their grammar from scratch. We shall hypothesize that the transmission of a grammar, whether through time or geographic space, is in general accompanied by equal or increased simplicity, and not by complication (reduction in generality).

It must be emphasized that all such statements are hypotheses about linguistic change; that is, they are statements that can be disproved. If a "counter-example" is found, which unequivocally demonstrates the incorrectness or implausibility of an hypothesis, then that hypothesis must be discarded or refined so as to be compatible with the data furnished by the counter-example. We cannot *prove* the hypothesis presented here (or any other hypotheses, for that matter). We can disprove them by finding a counter-example. We can support the hypothesis by showing that it is compatible with an ever widening field of data. We can show that an hypothesis in our theory is neutral with respect to a datum that is a counter-example to an hypothesis in a different theory of language change. But we cannot prove hypotheses in the way we prove, say, certain theorems in geometry given a set of undefined entities (point, line) and a set of axioms which state relations among these undefined entities.

4.1 CHANGES IN THE ADULT'S GRAMMAR

An hypothesis in our theory of language change is that changes in adult grammars are typically limited to minor alterations: addition of items to the lexicon, minor modifications in the formulation of a rule, addition of at most a few rules to a component of the grammar.

Note that "addition of a rule" does not mean that the adult looks around for a rule, finds one, and tacks it on to the end of his grammar in the way that a computer programmer might add to a previously written program an instruction to carry out an additional operation. The statement "Changes in adult grammars are typically limited to . . . addition of at most a few rules" is shorthand for the more complicated formulation: "Of the few ways a speaker's competence in his language changes once he has reached linguistic adulthood, one of the more common is most simply accounted for *in the linguist's model of this competence* as the addition of one or at most a few rules to the set of rules comprising a given component of the grammar. This rule acts on the previously produced output of the grammar and may modify it." Such a statement is neutral with respect to the internal makeup of the speaker's competence—the mass of brain cells, nerves, and so on, that account for speech in neural and physiological senses. In other words, what has happened to the inside of the adult speaker's head is something we at present haven't the faintest notion of. What we do know is that the grammar—that is, here, our account of the speaker's competence—registers such a change by having an extra rule added to it at some point.

What supports the hypothesis that adult grammars are rather severely limited in what can be done to them? One piece of the supporting data is the fact that people past a certain age find it next to impossible to learn a foreign language with native-speaker perfection. A child has no trouble learning his language, and he learns it perfectly unless physiological or emotional factors hinder him in some way. Children can even learn with native or near-native mastery two or more completely different languages if they are exposed in a natural way to speakers of these languages during their early years, say, before the age of thirteen or so (the exact age is subject to individual variation, but is likely near puberty). But adults do have all kinds of difficulty in learning a new language. The percentage of speakers handling two languages with native fluency, one of which they learned as an adult, must be vanishingly small in any culture.

The simplest explanation of this datum is that adults simply cannot do one of the things that children do without even being told: construct a complete grammar on the basis of exposure to a finite number of utterances in a language. Yet adults do exhibit changes in their speech performance. Even nonlinguists know that, and our theory of change must be compatible with the existence of adult change.

What are the typical changes that take place in adult grammar? One of the

easiest things an adult can learn, and one of the most trivial, is a new word. Before adulthood the number of new words a speaker learns each year must be in the neighborhood of a thousand or more, and up to complete senility and regression any adult in any language can learn a new word. Whether he habitually uses the word or not is irrelevant: as adults we can and do learn new words that we *can* use as long as we are in any measure open to experience. From the point of view of generative grammar, the learning of a new word is the addition of a new item to the lexicon—an utterly trivial alteration of the grammar. None of the existing output is affected unless the new word happens to displace an old one. No new grammatical rules must be learned to accomodate the addition. In short, addition of items to the lexicon—a typically adult thing to do—involves no major overhauling of the grammar; it is simply an addition to the lexicon, the formally most static component of the grammar.

A second characteristic kind of adult grammar change is the adoption, for some reason or other, of a prestige pronunciation. Such change can be very subtle and even independent of conscious effort by the speaker. Typically it occurs when the speaker wants to bring his speech more into line with the accepted standard or, and this amounts to the same thing, when he wants to be linguistically less striking in a particular social or geographic milieu. Consider, for example, the speaker from the South who goes to college in the North. In most cases, such a person would grow up saying [aˊ·ə̯] 'I', [aˊ·ə̯m] 'I'm', [kraˊ·ə̯m] 'crime', and so on. This is part of his "Southern drawl," and most speakers not from the South readily identify such a pronunciation as typically Southern. Our college student, if he is sensitive about such matters, might very well begin to pronounce the foregoing words closer to the norm of his new environment: [aj̯], [aj̯m], [kraj̯m].

How do we account for a change like this in the grammar of the college student? Whatever the exact way we do this, it should be clear that no major redoing of the grammar is necessitated. The principal phonological rules remain the same, not to mention the syntactic rules. In particular, the rules in the grammar of English which account for such alternations as *crime* : *criminal, finite* : *infinite, pronounce* : *pronunciation, profound* : *profundity* remain the same. Whether one says [kraj̯m] : [krɪminəl] or [kraˊ·ə̯m] : [krɪminəl], the simplest description in either case would have base forms with tense /ī/ in /krīm/ : /krīminæl/ and include in the grammar phonological rules laxing tense vowels such as /ī/ in certain environments, giving *criminal* with [i], and diphthongizing them in other environments giving *crime* with [ay]. The difference in the grammars of speakers of American English who say [aj̯] and those who say [aˊ·ə̯] is then the difference in a rule in the final, nonbinary section of the phonological component that prescribes the detailed phonetic description of [ay] from underlying /ī/. In the case at hand, probably the simplest way of accounting for the speech behavior of the transplanted Southerner is to assume that in his native grammar there is a low-level rule

[ay] > [a‹·ə̯]

and that he has added a rule which is ordered after this rule and gives

[a‹·ə̯] > [aɪ̯]

Whether or not this is the most desirable way of describing the present case, it is clear that no sweeping changes in the original grammar are needed. Almost all his previous rules remain intact after the Southerner learns his new pronunciation.

The claim that adults add late rules to their grammars is not devoid of empirical content. It implies that identical surface forms, even those deriving from distinct underlying sources, will be treated exactly alike by the rule added at the end of the grammar. This produces in certain cases the phenomenon known as "hypercorrection," which will be discussed further on in some detail. Thus, the claim that adults add late rules entails the prediction that adults, not children, tend typically toward hypercorrection, and this prediction is in fact borne out by the data on hypercorrection that are available.

Further cases of adult rule addition are those in which a rule spreads throughout an area. Such instances have been studied in depth by Labov (1963, 1965). We conclude from these studies that sound changes spreading within a few decades throughout large segments of the (adult) population represent instances of rule addition at a point relatively late in the grammar. One of the case studies concerns the variants found in New York City urban speech of the ingliding diphthongs [æə̯] and [ɔə̯]: cf. *bad dog* [bæə̯d dɔə̯g] : [bɛə̯d doə̯g] : [bɪə̯d dʊə̯g]. The occurrence of these variants shows definite age and ethnic correlations, and we assume that they represent in part at least the spread of a rule or rules throughout an adult population. In other words, they involve changes in the grammars of adults. In line with the hypothesis proposed earlier, we would expect that these vowel shifts in New York City are best described as instances of late rule addition that leaves intact underlying phonemic representations and the body of phonological rules that account for the considerable morphophonemic alternation in English. And, indeed, Labov states, "The far-reaching shifts and mergers observed in the long and ingliding vowel system of New York City . . . do not affect the morphophonemic system" (1965:102, n. 20).

A third kind of change often found in the language of adults, and not generally among young children, is hypercorrection. We shall see that hypercorrection is best understood as a sort of overlay of rules added to an already formed grammar, and not as a wholesale restructuring of the grammar.

The general sociological background of the phenomenon of hypercorrection is well known. It requires a situation in which certain items and casts of speech are recognized as prestige-bestowing. The speaker who hypercorrects desires to acquire the prestige conveyed by this sort of speech, has learned a

certain number of rules that bring his speech closer to the prestige norm, but applies these rules incorrectly, thus producing a hypercorrect form. A case in point is hypercorrect *whom* for *who*: *There you see the man whom we believe is the murderer*; *Whom do you think will be the next President? Who* is correct in such contexts, and the hypercorrect use of *whom* obviously has something to do with the fact that the vast majority of speakers of English "incorrectly" use *who* for *whom* in, for example, *Who do you see?*; *Is he the man who you wanted to talk to?* The hypercorrecting speaker adds a transformational rule that replaces *who* → *whom* in certain surface syntactic contexts. When applied to the output that his grammar has already produced, this rule will give correct results in *Whom do you see?* but hypercorrection in *Whom do you think will be the next President?* This particular instance of hypercorrection has been thoroughly discussed by Klima (1965), where it is shown that the grammar of the hypercorrect *whom* speaker has an extra transformational rule added at the end of a group of related transformations. There has not been any major change in the grammar; in fact, to use *who* and *whom* in the way sanctioned by Miss Fidditch the speaker would have to reorder two transformations in his grammar, and this the hypercorrect speaker has not done.

A case of phonological hypercorrection from Low German further illustrates the superficiality of the change undergone by the grammar of a speaker who produces something hypercorrect. Low German, originally widely spoken in Northern Germany, has little prestige nowadays. No one speaks "pure Low German" except possibly in rural areas far removed from the inroads of modernity. For any number of historical reasons Low German has retreated before Standard High German, so that today one can hardly rise socially in Germany without High German. Low German and High German are closely related, sharing a large number of cognate lexical items such as Low German *ik* : High German *ich* 'I'. The chief identifying parameters of Low German are phonological ones such as Low German stops corresponding to High German affricates or fricatives, as in *ik* [ɪk] versus *ich* [ɪç]. The major correspondences are: Low German [d] = High German [t]; Low German [p t] = High German [pf ts] in certain environments and [p t k]= [f s x] in others; Low German [i: ü: u:] = High German [aɪ̯ ɔɪ̯ aʊ̯]. Details vary from dialect to dialect; the following examples are from a dialect of Low German spoken along the lower Elbe (Keller 1961:339–379):

Low German	*High German*	*Gloss*
dɔxtər	tɔxtər	daughter
pɛpər	pfɛfər	pepper
ti:t	tsaɪ̯t	time
makən	maxən	to make
bɛtər	bɛsər	better
bli:bən	blaɪ̯bən	to stay
fü:st	fɔɪ̯stə	fists
bu:k	baʊ̯x	stomach

Knowing that Low German has little prestige, that it is similar to High German in many ways, we expect and find a good deal of hypercorrection in the speech of native Low German speakers striving to improve their position and chances in a High German society. We find, for example, the hypercorrect form *[baxən] 'to bake' alongside correct Low and High German [bakən]. The explanation is clear. The native Low German speaker, accustomed to hearing High German [x] or [ç] (after front vowels) in place of his [k] in many items such as [makən] : [maxən] 'to make' and [ɪk] : [ɪç] 'I', adds to his grammar a rule shifting $k > x$ (and $p > f$, $t > s$) in the appropriate environment (primarily post-vocalic). In instances like those of 'to make' and 'I', the result is fine: our Low German speaker trying to speak High German comes up with a form not appreciably different from that of prestigious High German.

The catch leading to hypercorrection is that Low German [k] does not always correspond to High German [x] in the requisite environments; some postvocalic [k]'s in Low German correspond to High German [k]'s: compare Low German and High German [bakən] 'to bake', Low German [ki:kən] versus High German [kʊkən] 'to look'. The correct form produced in the Low German grammar is [bakən], and the correction rule $k > x$ is erroneously applied, yielding *[baxən].

Similar statements could be made about other sets of correspondences. For example, Low German [u:] corresponds to High German [aʊ̯] in, say, the word for 'stomach', but there are also cases of Low German [u:] = High German [u:]. We have this in Low German [ju:d(ə)] : High German [ju:də] 'Jew', which moreover is a case of Low German [d] = High German [d] as against the also occurring correspondence [d] : [t]. The added rules of hypercorrection when applied to correct Low German [ju:d(ə)] produce hypercorrect High German *[jaʊ̯tə].

Hermann (1931:37) quotes an amusing story as an example of hypercorrection: The story was told by the prominent Low German writer Rudolf Kinau concerning his first day in school, where naturally it was expected of him that he should speak only High German in class. Kinau's nickname was *Rudel*, so that in his native Low German he would have pronounced his name [ru:dəl ki:naʊ̯]. This is also a perfectly acceptable High German name, and in High German it would also be [ru:dəl ki:naʊ̯]. When asked on that first day to give his name, in High German of course, Kinau promptly replied *Rautzel Keinau* *[raʊ̯tsəl kaɪ̯naʊ̯].

Part of this is readily understandable in view of what has already been said about German hypercorrection by native Low Germans. The rule [i: ü: u:] > [aɪ̯ ɔɪ̯ aʊ̯] has been added on the Low German grammar at a low level, and it is erroneously applied to the [i:] and [u:] in [ru:dəl ki:naʊ̯] to give *[raʊ̯tsəl kaɪ̯naʊ̯]. The [ts] in *[raʊ̯tsəl] is a sort of hyper-hypercorrection, reflected in the grammar by a reversal in the order of application of two rules. It was earlier pointed out that Low German [d] corresponds in many cases to High German [t] (*Dach* versus *Tag* 'day'), and that Low German [t] corresponds to High German [ts] (*to* versus *zu* [tsu:] 'to'). To get from Low German to High German in the simplest way, two rules must be added to the Low German grammar:

A. $t > ts$

B. $d > t$

The order of application of the two rules must be A followed by B. These rules, applying to what has already been produced in Low German, will give the right High German forms:

Low German Forms:	fadər 'father'	taɪ̯n 'ten'
Rule A:	.....	tsaɪ̯n
Rule B:	fatər	.....

We get [fatər] and [tsaɪ̯n], which differ only in relatively minor details of vocalism from the correct High German forms [fa:tər] and [tse:n].

What Rudolf Kinau did on his first day of school, in trying under pressure to render his name in High German, was to reverse the order of application of the two rules, at least for the one time in question: the [d] in *Rudel* was changed by Rule B to [t], which then qualifies as input to Rule A and undergoes [t] > [ts]. The net result is hypercorrect *[raʊ̯tsəl kaɪ̯naʊ̯] from correct [ru:dəl ki:naʊ̯]. One cannot assume that Kinau consistently applied the rules in the order B followed by A, which is the opposite of correct, for no German speaker could go around for long saying *[fatsər] for correct [fa:tər] 'father' and *[tsax] for correct [ta:k] 'day' without having his mistake firmly brought home to him in some way or other.

The point of the example is that rules of hypercorrection do not seem to be firmly embedded in a natively acquired grammar. In hypercorrection there is no perfect mastery of the correct grammar and sequence of rules, no restructuring of the kind that would proceed from native internalization of the grammar. Rules of hypercorrection seem rather to be an inorganic and somewhat ephemeral superstructure built onto the firmer foundation of a grammar acquired through the normal process of language acquisition.

4.2 GRAMMAR CONSTRUCTION IN THE CHILD

When we leave the adult's grammar and turn to the child's grammar and way of acquiring language, we enter a field bristling with question marks. Enough is known, however, to suggest certain hypotheses about language change.

How does a child learn to speak his language? As little as is known about

this engrossing and complex subject, certain notions seem definitely wrong. One is the idea that the child, like a parrot learning to talk, repeats what is said to him and in his presence. The older the child becomes, the more he has heard; consequently he is able to say more and his imitations approach the adult model.

The most telling argument against this gross version of the imitative theory is the child's ability to be linguistically creative even at an early age. A normal child of three can say all kinds of things which he could not possibly have heard. He has acquired, in other words, the ability that all native speakers of any language possess: the potential of constant and, in principle, infinite inventiveness. In the course of a single two-hour period a child at age three and a half produced the following monologues while talking to himself at bedtime:

A. (1) cat (many times)
(2) two (many times)
(3) bats
(4) the cat sees two bats

B. (1) pig (many times)
(2) big
(3) sleep
(4) big pig sleep now

Repetition drills like these are quite common among children (Weir 1962). What is interesting is that the end results of monologues A and B—*the cat sees two bats* and *big pig sleep now*—were both utterances that the child in question had never heard before either in conversation or in stories read aloud. For him they were new and unique creations.

Whatever the exact nature of the child's competence and whatever characteristics of the human being enable him to arrive at a competence in so short a time, it is certain that a child soon goes beyond the corpus of utterances to which he has been exposed. Following McNeill (1966:19) we propose to explain this by assuming that child speech is not garbled output of a complete adult-type grammar but the product of a first, relatively simple grammar. In this view of child language the "telegraphic speech" (Brown and Fraser 1963) so characteristic of young children (*who you?*; *no more*; *want water*) is not so much a falling short of adult models as the output of a grammar—a system of internalized rules—that is shorter and simpler than that of adults. The progress toward adult speech is a constant process of readjusting and adding to the previous grammar.

The child acquires language roughly as follows. He is provided with a small body of data about his language—the corpus of utterances he has heard. On the basis of this limited corpus he constructs a grammar that produces a set of utterances approximating adult speech. This first grammar is then a sort of

hypothesis about the native language from the child's point of view: if I carry out certain rules I will produce my native language, as it were. In general this hypothesis will be at least partially incorrect at first. Its incorrectness will become apparent to the child in various ways: his mother or siblings laugh at his attempts or correct them; he doesn't get what he thought he was asking for; certain internal comparisons, which we know little about, may cause him to reject certain constructions. The child then refines his hypothesis—his grammar—perhaps extending its range in light of the larger number of utterances he has heard, and again produces utterances. By frequent and constant hypothesis testing of this sort the child eventually arrives at a grammar of his language, a competence that underlies speech output close to acceptable adult speech.

Note that we must not take this description of grammar building too literally. What we mean by "a child constructs his grammar" is not that he builds something the same way he builds a house out of Lincoln Logs or a ferris wheel with an erector set. By the process of hypothesizing, testing, evaluating, repeated over and over again, the child develops a competence in his language that we as linguists may represent as a set of components containing ordered rules.

This process of grammar construction in the child is unique—unique as an idiolect, as unique as a personality. Yet like an idiolect or a personality, the grammar a child constructs is not totally different from those of the speakers who cared for him, played with him, and read to him during the years of language formation, just as a child's personality is never completely different from the personalities of the people who move about in his world.

The parallel between grammar and personality is not a bad one. A child developing his personality constructs hypotheses about behavior by observing the social and emotional performance of those around him. He tests these hypotheses in various ways, and they are verified, or rejected in equally various ways. He then discards, corrects, or adds to his previous hypotheses about his place and role in the world, tests these, and so it goes. Unless some severe emotional disturbance is created which disrupts the process of personality building completely (as in autism), the child ends up with a well defined personality—a system of rules determining his image of himself in its social setting—and this basic personality tends to remain with him throughout life. His personality has much in common with those of his overseers and playmates, yet differs from them in various ways both good and bad.

A child's grammar building is to a certain extent analogous to his personality building, yet is typically less of an original product than his personality; for one thing, the child's grammar resembles adult grammars more than is usually the case with personality. If this were not the case, then there would have long since been a pressing need for "linguotherapists" whose relation to their subject's language would parallel a psychotherapist's relation to his patient's psyche. But the instructive point of similarity is that in both

cases the child constructs something uniquely his yet not radically different from the models present in his environment. The potentialities of deviation are great in any such original process.

Pressing the analogy one step further, we point out one crucial feature in which grammar construction and personality construction tend to diverge radically. A child rarely, if ever, constructs a grammar more complex than that of his models. He will accept his grammar as it is with perhaps minor deviations, or he will simplify it in various ways and for various reasons; but he will not complicate it. This is not a general rule of personality construction. There are many cases in which the child constructs a more complex, a more maze-like picture of himself and his place in the world than any of his models has constructed for himself.

This sketch of the child's grammar acquisition highlights one of the chief sources of linguistic change: the transmission of language to the new generation, or, to use a less misleading figure of speech, the acquisition of language by each child in a new generation. We saw in Section 4.1 that adult grammars do not seem susceptible to much change once they have "congealed." Yet the histories of languages are replete with examples of radical changes that cannot be reasonably accounted for by rule addition such as might occur in adult grammars. For example, we assume that rule reordering as discussed in Section 3.3 occurs in the child's learning of a language and not within the grammar of an adult, and we attribute such changes to *grammar simplification* (optimization).

The role of grammar simplification is basic to our conception of linguistic change. The underlying idea was originally stated by Halle (1962:64):

> The ability to master a language like a native, which children possess to an extraordinary degree, is almost completely lacking in the adult. I propose to explain this as being due to deterioration or loss in the adult of the ability to construct optimal (simplest) grammars on the basis of a restricted corpus of examples. The language of the adult—and hence also the grammar that he has internalized—need not, however, remain static: it can and does, in fact, change. I conjecture that changes in later life are restricted to the addition of a few rules in the grammar and the elimination of rules and hence a wholesale restructuring of his grammar is beyond the capabilities of the average adult.

Note carefully what this does *not* say. It does not say that change is confined to the child. It does not say that children can speak only with flawless or maximally simple grammars in their heads. It does say that children can and often do construct a grammar formally simpler than adult grammars. It does imply that a child, creating a grammar from the finite and fairly small corpus of examples he has to go on, can come up with a competence—an internalized grammar—that is simpler than an adult grammar yet underlies a speech

output either identical with adult speech for all practical purposes or different in relatively minor ways.

Let us examine some of the evidence that supports the notion of child optimization of grammar. That children simplify is well known and obvious to everyone. This is particularly evident at the morphological level: probably every English-speaking child has at some time said something on the order of *I goed*; *two foots*. This simplification is the extension of a pattern of inflection from the regular cases to irregular ones. Ervin (1964) has called our attention to an interesting progression of simplification. The first verb forms which are used in child speech are unmarked for tense: *Where the car go?*; *What he do?*, where the context alone signals which tense is meant. The first past tense forms that emerged in Ervin's observations were irregular ones, such as *went*, *did*, *came*, and these at first were formed correctly. This is not surprising since these have a high frequency of occurrence and a child will hear them only in their correct form in a family speaking standard English. Soon, however, the children observed by Ervin went over to regular past tense formations such as *goed*, *doed*, *comed* even at a time when the child had control of only a few weak verbs that could serve as a model for the extension. In fact, in Ervin's tape-recorded sample the children produced no weak verbs, though this is attributed to the relative infrequency of weak verbs in the speech normally directed toward a child. In other words, almost as soon as he got the chance the child substituted incorrect but regular past tense forms for his earlier correct but irregular forms. From the point of view of the child's grammar, he has added a rule

$$\text{Verb} + \text{Past} \rightarrow \text{Verb} + /\text{d}/$$

which then functions for all verbs irrespective of their idiosyncracities in adult speech.

In the same way, a child's first acquired syntactic rules tend to be simplified versions of similar rules in adult grammars. One of the first rules of syntactic formation internalized by the child has the form (McNeill 1966:23)

$$S \rightarrow (P)\frown O$$

where *P* and *O* stand for "pivot class" and "object class" respectively (Braine 1963). This rule produces expansions of *S* as either *O* or $P\frown O$, giving a set of one and two word sentences: *allgone milk*; *byebye boat*; *shoe*; *hot*; *baby*; *this baby*. In fact, much of the "telegraphic" nature of child speech results directly from simplified versions of rules like this, and it has even been suggested (McNeill 1966:19) that children produce telegraphic speech for much the same reasons that adults send telegrams saying: BROKE. NEED MONEY instead of I'M BROKE AND I NEED MONEY. The reason is cost—either financial or cognitive. As the adult omits unnecessary words to make his

message cheaper, the child economizes on the amount of mental effort needed to express what he wants.

Many cases of simplification in the child's acquisition of language are more subtle than either of these examples from morphology and syntax. The following instance is based on personal observations by the author. Most kinds of American English have a well-known phonological rule that makes *t* into *d* or into a half-voiced (or voiced) alveolar flap ("voiced *t*") after a stressed vowel and before an unstressed vocalic segment. In these dialects *write* : *writer* is [raɪt] : [raɪ̬dɚr], and *latter* differs from *ladder* at most in slight extra vowel length in *ladder*, both being approximately [lædɚr] ([d] here is a cover symbol for the voiced alveolar stop [d] or the half-voiced alveolar flap [ɾ]).

A child, closely observed in his linguistic behavior from the age of two and a half on, grew up in a family and in a region (Madison, Wisconsin) where the *t*-voicing rule was operative. Between the ages of five and six, however, he was observed to pronounce, for example, *fight* : *fighting* as [faɪ̬tʰ] : [faɪ̬tʰɪŋ] and also *fighter* [faɪ̬tʰɚr], similarly *sit* : *sitting* [sɪtʰ] : [sɪtʰɪŋ]. Now at that time and earlier no one in his environment or in the television programs he watched said such things: they all had [sɪtʰ] ~ [sɪt] : [sɪdɪŋ], and so on. This is best explained as simplification: his grammar was not appreciably different from that of his adult models, but simpler in that it was a rule shorter, for it lacked the *t*-voicing rule present in the adult grammar. Both the child and the adults had underlying /sit/ : /siting/ for *sit* : *sitting* since the adults have a *t/d* alternation. The *t*-voicing rule in the adult grammar would give the biallomorphic forms [sɪtʰ] : [sɪdɪŋ]. The child, having no such rule, would say [sɪtʰ] : [sɪtʰɪŋ] with a single allomorph [sɪtʰ].

Note too that the child said *water* [wadɚr] and *potty* [pʰadiy] as did his elders. The [d]'s in these words were invariant in the speech of the adults; that is, they underwent no phonological alternation with [t]'s elsewhere in the paradigm. Hence the child had no reason to posit base forms with /t/ in such forms. (He had not heard *pot* at this point, only *potty*.) *Water* and *potty* would then have the base forms /wadr/ and /padi/, which give phonetic [wadɚr] and [pʰadiy]. In *sit* : *sitting* and similar alternations, in which the adults around the child presented him only with [sɪtʰ] : [sɪdɪŋ], the child constructed base forms with /t/ giving /sit/ : /siting/ and ignored the [d] in *sitting* by not adding a *t*-voicing rule to his grammar.

We can call this either *imperfect learning* (Kiparsky 1965) or *grammar simplification*. The child has obviously not quite arrived yet at the grammar of the adults around him; that is, he has learned his language imperfectly. His grammar is simpler by a rule than that of his parents and baby-sitters, and this simplification shows up in his speech output as reduction in the amount of allomorphic variation in *fight* : *fighting*, *beat* : *beating*, and so on. At about the age of seven, this child finally added the rule of *t*-voicing to his grammar and started saying [biːt] : [biːdɪŋ], [sɪt] : [sɪdɪŋ], and so on.

This example invites comparison with the case discussed in Section 3.3 of rule loss in Yiddish. There, a rule for devoicing terminal obstruents was lost, producing *veg* : *vege* and *tog* : *teg* from earlier surface forms *vek* : *vege* 'path, paths' and *tak* : *tage* 'day, days'. One of the crucial bits of evidence of rule loss was the presence in contemporary Yiddish of relic forms like *avek* 'away' from original *veg* 'path', indicating the previous existence of a rule of terminal devoicing.

Suppose now that the child discussed in regard to *t*-voicing had retained into adulthood his grammar of age five. Suppose further that other children of the same and following generations also retained a grammar of English that lacked the *t*-voicing rule. This situation is not completely far-fetched because such a grammar is simpler than one containing the *t*-voicing rule and because there are dialects of English (like British English) without the rule. We would then reach a point in several generations where a sizeable portion of the population would be saying [sɪtʰ] : [sɪtʰɪŋ], [raɪ̬tʰ] : [raɪ̬tʰɪŋ] (*write* : *writing*). From the viewpoint of historical linguistics we would know this to be a case of rule loss for the same kind of reason as in Yiddish: the existence of "relic" forms [wadɚr] *water*, [bɪdɚr] *bitter*, [lædɚr] *latter*. If we had sufficient knowledge of the history of English and other dialects of English, we would know that such forms originally had *t* in them, and the fact that they now show [d] would point back to a stage when the *t*-voicing rule was operative.

Note that rule loss might better be termed "rule nonacquisition" to emphasize the likely mechanism by which rules are lost from a grammar. However, the notion of rule loss has been in historical linguistics for a long time, and it is preferable to retain the traditional terminology for this kind of primary change.

Examples such as these, which can be multiplied by close observation of child speech, support the proposal that children simplify (optimize) the grammar that they construct. This does not mean, of course, that they must always simplify or that they can never acquire more difficult grammar rules. The maturation process in child language is precisely characterized by the acquisition of additional rules, the refinement of already acquired rules—in general the construction of a larger and more complex grammar. But in being presented with the data of his language, each child draws his own conclusions about what kind of grammar has produced the data. Each child in each new generation takes a fresh look at the situation, as it were, and the result is often simplification of a sort beyond the capabilities of adults, who have completed the construction of *their* grammars—at the least beyond the capabilities of the average, linguistically unsophisticated adult. The restriction on restructuring in adult grammars may not be quite so severe for adults who, for one reason or another, have a greater than average concern with language. Anyone who (like the present author) didn't get *who* and *whom* straight until the first year of college has apparently succeeded in reordering a pair of transformational rules. This is what the Klima (1965) analysis would suggest.

At any rate, the assumption is that the average person, as opposed to the linguist or even the linguistically astute educated person, cannot change his grammar in radical ways once linguistic adulthood is reached.

This view, which attributes to the new generation a considerable participation in linguistic change, is by no means novel or revolutionary. It was quite widely held among linguists in the late nineteenth century. Hermann Paul, writing around 1880, stated flatly: "The chief cause of sound change lies in the transmission of sounds to new individuals" (1960:63). Rousselot, in the last decade of the nineteenth century, concluded his immensely detailed study of sound changes in a group of French dialects with the observation: "The principle [of linguistic evolution] resides in the child. . . . Parents set the stage for [linguistic] evolution; but the real impetus for this evolution comes only when the children enter into possession of their language" (1892:412–413). Paul Passy argued cogently that many phonological changes arise in the child's acquisition of speech: "All the major changes in pronunciation that we have been able to investigate originate in child speech" (1891:231). And, most eloquently of all, William Dwight Whitney said of the continuity of language through generations:

> Human institutions in general go down from generation to generation by a process of transmission like that of language, and they are modified as they go. . . . No one has ever yet been able to prevent what passes from mouth to ear from getting altered on the way. . . . Although the child in his first stage of learning is more than satisfied to take what is set before him and use it as best he can, . . . the case does not always continue thus with him; by and by his mind has grown up . . . and begins to exhibit its native and surplus force; . . . it modifies a little of its inherited instrument, in order to adapt this better to its own purposes (1883:34–35).

One must note carefully what the arguments of these last two sections imply, and what they do not imply. The principal point is that the potential for change is more severely constrained in adult grammars than in child grammars. What changes occur in adult grammars seem to be few and minor, mostly limited to addition (and, occasionally perhaps, loss) of late rules. The child, in constructing his grammar, obeys less rigid constraints: he can reorder rules, lose rules, generalize rules—in short, can change his grammar in all sorts of ways.

Such a view, neither denying change in the adult nor confining it to children, seeks to state more precisely what types of changes can occur in child and adult grammars. The hypotheses here and in the preceding section have, a priori, a good deal going for them, as has been pointed out, but it is not claimed that every detail of the arguments presented here will remain unchanged under lengthy investigation. No counter-evidence is known to the present author, but then there have been few empirical studies explicitly of what is possible in adult grammars and what is not. Nor have there been enough investigations of the child's acquisition of grammar to enable us to make much more than reasonable hypotheses about this process. These are serious gaps in the data of linguistics; and further research into linguistic change will depend greatly on how well and how completely these gaps are filled.

In particular, little is known about the constraints on change in a child's grammar: given a particular grammar, which rules can be lost (not acquired), which pairs of rules are especially suitable candidates for reordering, which rules are likely to be simplified? Probably such changes usually can affect only later rules in the phonology; the *t*-voicing rule, discussed earlier as an instance of rule loss, is a late rule in English phonology whose effect on the phonetic output is relatively minor. It is a little difficult to imagine the loss of an early rule that radically changes the forms it acts on, of rules such as Velar Softening or Main Stress in English phonology (see Chomsky and Halle 1968:239–240).

Finally, the claim that adult change is largely confined to rule addition does not deny innovation (rule addition) in the child. It is probable that rules of assimilation often arise in the child's acquisition of grammar. Since assimilation produces ease of articulation in some sense, we expect children to add just such rules. Doubtless this tendency towards ease of articulation accounts for the widespread occurrence of certain rules in approximately the same form in different languages, for example, in the rule of nasal assimilation (nasals agree with the following consonant in point of articulation) found in so many of the world's languages. Even rules less obviously assimilatory may arise in child grammar construction. The author has observed several cases of English-speaking children, from ages two and a half to four, who incorporated in their grammars an optional rule devoicing word-final obstruents; they said, for example, [muwf] 'move', [dɔk] 'dog', and so on. These children did not keep these rules beyond the age of five.

So children seem to go beyond simplification in the ordinary sense to innovation. Likewise, child innovations probably are often (if not always) assimilatory in nature; that is, they are simplifications in a "local" sense. As certain as it is that child innovation occurs, at present little is known about which phonological rules arise via child innovation.

4.3 A MODEL OF LINGUISTIC CHANGE

Our conception of language transmission from parent to child can be summarized as follows. The parent has a competence, an internalized crammar, underlying his speech output. Though the grammar of the adult gannot undergo a radical transformation, it is susceptible to innovation in the

form of rule additions and minor rule changes. The child, developing his grammar from the speech output of his parents and older peers, arrives at a linguistic competence not radically different from that of the adult. The child's competence reflects not only the original grammar of the adult but also those innovations that the adult grammar may have undergone. The child will optimize—simplify—and in the process linguistic change may result.

Let us consider two hypothetical examples. Assume a language with the stops /b d g p t k/, and suppose that adults add a rule merging *b d g* > *p t k* unconditionally. Children, hearing only [p t k], would have no reason to posit underlying /b d g/; hence they would construct a grammar (the optimal grammar) containing only underlying /p t k/ and no rule devoicing [b d g]. If, however, the context-sensitive rule *b d g* > *p t k* / —— # had been added by the adults and if this rule produced morphophonemic alternations such as *haba*:*hap*, *sidu*:*sit*, *pego*:*pek*, then no simpler grammar would account for the data. (Compare the case of terminal devoicing in German discussed in Section 2.2.) In this case, the child will incorporate this rule in his grammar.

Let us now consider some actual cases. In many dialects of English initial [hw] has been reduced to [w]: *whip*, *what*, *when* are pronounced [wɪp], [wʌt], [wɛn], not [hwɪp], [hwʌt], [hwɛn]. The change *hw* > *w* has occurred throughout the entire British Isles except for the northernmost counties of England; in an extensive coastal section of the Middle Atlantic States, according to the records of the Linguistic Atlas of New England (Kurath and McDavid 1961:178); and, by informal observation, in other large sections of the United States such as the Middle West. (See Section 5.1 for further discussion of this change and for more detail about its phonetics.)

Let us now imagine ourselves in a time when the [hw] pronunciation by assumption was still universal. How can the sound change have taken place and spread? We assume that someone, for some reason, quit saying [hwɪp], [hwʌt], [hwɛn] and began to say [wɪp], [wʌt], [wɛn]. In our account of this speaker's competence we add a rule to his grammar:

$$4.1 \quad \begin{bmatrix} -\text{vocalic} \\ -\text{consonantal} \end{bmatrix} \rightarrow \emptyset \; / \; \text{——} \; \begin{bmatrix} -\text{vocalic} \\ -\text{consonantal} \\ +\text{back} \end{bmatrix}$$

(A semivowel, e.g. [h], is deleted before the semivowel [w].)

We can, if we like, speculate on *why* this rule was added. Perhaps the speaker thought *w* sounded better than *hw*, perhaps *hw* was harder to pronounce than *w*. Such speculation is interesting but outside our immediate major concern, which is to give an account in our grammar of a change in speech habits. The simplest way to do this is to assume that our speaker has added Rule 4.1 to his grammar. This represents an innovation in the speaker's grammar, something new with him.

As before the rule addition, the speaker will have lexical entries with /hw/: /hwip/, /hwat/, /hwen/. Rule 4.1 operates on these lexical items, deletes the /h/, and gives phonetic output with initial [w]. It is altogether possible and likely, in view of the large amount of British territory in which *hw* > *w*, that Rule 4.1 spread among adult speakers of English. If this is true, then Rule 4.1 was added to the grammars of these speakers too as an innovation, presumably with [wɪp] retaining the underlying form /hwip/.

Now we come to the child learning the language. From speakers who have Rule 4.1 in their grammars, the child will hear only forms with [w]: [wɪp], [wʌt], [wɛn], [wɛðir] *whether* identical with *weather*. He has no reason to assume underlying /hw/ in such words, so that the *child* enters in the lexicon of *his* grammar the underlying forms /wip/, /wat/, /wen/, and so on. There are no *hw* : *w* alternations in the language of his parents to motivate the inclusion of Rule 4.1. Thus, the simplest grammar is one containing no Rule 4.1 and no underlying /hw/ forms in the lexicon. The child thus arrives at a grammar which produces (in this one respect) the identical output as the parent grammar and which is simpler. The output of this grammar in turn serves as the primary data for the language acquisition of the next generation, whose grammar will likewise lack underlying /hw/ and Rule 4.1. In this way we conceive of the change as having spread first as an innovation in the grammars of adult speakers and then as a simplification of the next generation's grammar.

In arriving at the lexical entries with /w/ replacing /hw/, the grammar has undergone simple *restructuring*. We define restructuring as any change in underlying representations. Thus, the four types of primary change discussed in Chapter 3, since none of them necessarily requires change in underlying (systematic phonemic) representations, are not restructuring. It should be noted that usage differs concerning the term *restructuring*. Other linguists distinguish two categories of change: innovation and restructuring. In this usage restructuring comprises rule loss and reordering, simplification, and change in underlying representations.

In the preceding two sections it was argued that adult grammar change was confined to rule addition. Rule loss and reordering, simplification, and restructuring originate in the child. This is the puristic picture. It may well be that adults are capable of participating in certain minor grammar changes other than rule addition, e.g. loss or simplification of certain low-level rules. Adults may even be capable of minor restructuring, though we assume subject to disconfirmation that major change in underlying representations is beyond the adult's ability. Optional, stylistic rules in adult grammars support this assumption: adults use them or not at will, and the lexicon continues unchanged throughout.

This example of innovation in the adult grammar followed by restructuring in the child grammar could, it should be pointed out, be explained in a somewhat different way. Since forms in *w* and *hw* do not alternate phonologically,

an adult might have restructured his lexicon vis-à-vis these morphemes. The restructuring involved here would be minor, and it might very well be within an adult's capabilities to alter his grammar to this extent.

Innovation does not always lead to restructuring in the subsequent generation. If no simpler grammar produces the same speech output, there will be no restructuring. The Great Vowel Shift in English is a case of this kind (Chomsky and Halle, 1968:249–289).

Middle English (spoken approximately from 1100 to 1500) had the tense vowel system:

ī ū

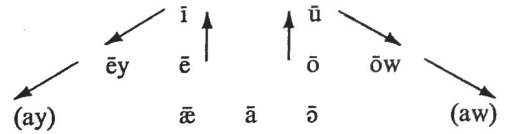
ē ō

æ ā ɔ

By the Great Vowel Shift we understand the set of changes in which /ī ū/ became diphthongized to [ay aw] by way of [e:y o:w] and /ē ō/ were raised to [i: u:], for example:

Middle English	*Modern English*
mīn	mine [mayn]
þūsend	thousand [θawzɨnd]
sēk	seek [si:k]
schō	shoe [šu:]

We can represent these changes diagrammatically as follows:



To account for just this part of the data, we assume two innovations in the grammar of Middle English around 1500. (We ignore here changes affecting the low vowels.) The first of these is a diphthongization rule affecting *ī* and *ū*:

$$4.2\ \emptyset \rightarrow \begin{bmatrix} -\text{vocalic} \\ -\text{consonantal} \\ \alpha\ \text{back} \end{bmatrix} / \begin{bmatrix} +\text{vocalic} \\ -\text{consonantal} \\ +\text{tense} \\ +\text{high} \\ \alpha\ \text{back} \end{bmatrix} \text{——}$$

(*ī* > *īy* and *ū* > *ūw*. This rule inserts the glide [y] or [w] depending on whether the preceding vowel is *ī* or *ū*.)

The second innovation, added to the grammar after Rule 4.2, is the Vowel Shift rule proper:

$$4.3\ \begin{bmatrix} \alpha\,\text{high} \\ -\,\text{low} \end{bmatrix} \rightarrow [-\alpha\,\text{high}] / \begin{bmatrix} +\,\text{tense} \\ +\,\text{stress} \end{bmatrix}$$

(Tense, stressed, nonlow segments exchange their highness values: *īy* > *ēy*, *ūw* > *ōw*, *ē* > *ī*, *ō* > *ū*.)

Derivations then go as follows:

Underlying:	ī	ē	ō	ū
Rule 4.2:	īy	...	...	ūw
Rule 4.3:	ēy	ī	ū	ōw

But this tells only part of the story. Middle English had a rule that laxed vowels before consonant clusters and when followed by two syllables the first of which was unstressed. The approximate form of this rule was:

$$V \rightarrow [-\text{tense}] / \text{——}\ C \left\{ \begin{matrix} C \\ \begin{bmatrix} V \\ -\,\text{stress} \end{bmatrix} CV \end{matrix} \right\}$$

The result of these rules—the laxing rule together with Rules 4.2 and 4.3—was to produce phonological alternations in Early Modern English of the types:

Early Mod. Eng.	*British Modern Eng.*	*Examples*
[ey] : [i]	[ay] : [ɪ]	crime : criminal
[ow] : [u]	[aw] : [ʌ]	profound : profundity
[i:] : [e]	[i:] : [ɛ]	keep : kept
[u:] : [o]	[u:] : [ɔ]	goose : gosling

Here, we are ostensibly faced with the same situation as before in regard to the reduction of *hw* > *w* in the adult grammar. Rules have been added to the adult grammar. The underlying forms have remained the same (e.g. the underlying forms of *keep*:*kept* [ki:p]:[kept] are /kēp/:/kēpt/ as in pre-Vowel Shift Middle English); only the phonetic outputs are different: the speech has changed but the lexicon and its representation have not.

But unlike the case of the child constructing a grammar on the basis of speech produced by his elders with the innovation *hw* > *w*, here there is no simpler grammar that will account for the same output. Because of phonological (morphophonemic) alternations, the simplest grammar still contains underlying tense vowels in *crime*, *criminal*, *profound*, *profundity*, *keep*, *kept*, and so on, and it still contains Rules 4.2 and 4.3 (as well as the laxing rule) in essentially the same form. These rules are still present in the grammar of Modern English and have been for the past four centuries or so. This is to say that English has undergone little restructuring among tense vowels

during this time. Its underlying phonological representations of forms in tense vowels have changed but little since Middle English, and the synchronic analogues of the Great Vowel Shift as well as other historical innovations have been passed along from generation to generation in approximately the same form.

Thus, there are different modes of simplification in the child generation. The adult may have added a rule giving him a nonoptimal grammar; the child will construct an optimal grammar producing the same output. If there is no simpler grammar that produces the same output as that of the adult grammar plus the innovation, the child's grammar can consist of the adult grammar plus innovation. In these two cases the child's grammar output—his speech—will not differ from that of the adult. On the other hand, if the child goes further and simplifies by losing or generalizing a rule, thus constructing a grammar simpler than the adult optimal one, then his speech will differ correspondingly from adult speech.

Before considering additional cases of diachronic change, perhaps it would be well to examine Figure 4.1, the schematic representation of the process of linguistic change, which is based on Klima (1965:83).

LAD stands for Language Acquisition Device, which is a "black box" construct designed to cover the child's whole complex process of receiving the primary data of his language and developing from it the optimal (descriptively adequate) grammar for his language. Thus, Generation 2 utilizes the Speech Output of Generation 1 to arrive at an Optimal Grammar. In the course of adult life Innovations may be added on to this grammar, giving what we have called the Adult Grammar of Generation 2. The Speech Output B of Generation 2 then serves as input to the LAD for Generation 3, and on it goes.

There are several points in the theory underlying this representation of diachronic change that merit special comment. First, our model does not represent *speech* as changing into *speech* with time. Speech output at one stage is not mapped directly into later speech output: no arrows connect the Speech Output at a given stage with the Speech Output at a different stage. What does change is the grammar vis-à-vis different stages. The grammar of one stage is developed on the basis of speech produced by a grammar at an earlier stage, and the grammar of a speaker may undergo innovations—rule additions. To use the terminology developed in Chapter 2, linguistic change is change in *competence*, not change in *performance*; it is change in the grammar, not originally change in the output of that grammar. This conception of change is all important, as we shall see in Chapter 5, where the nature of phonological change is probed in detail.

A second point is the matter of comparing different stages of the same language. What does it mean to say that Middle English and Modern English are different stages of the same language? What's the same? This is something of a classical antinomy in linguistics, whose synchronic

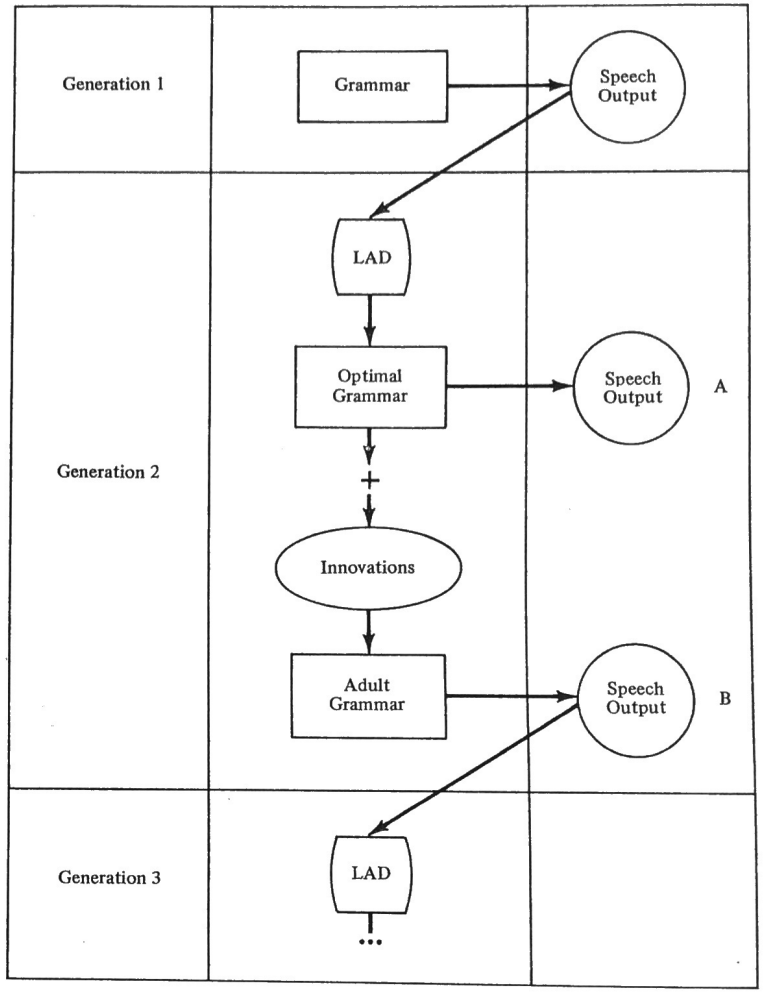


FIGURE 4.1

A MODEL OF LINGUISTIC CHANGE

counterpart was discussed in Section 3.1: what it means to say that linguistic systems A and B are dialects of the same language. If we accept the Saussurean dictum that linguistic elements are defined synchronically by all other elements in the system at that instant in time, then it is not obvious how we can speak of "correspondences" between elements in the language at different times (cf. Hoenigswald 1960:27f). In what sense does Modern English /ay/ correspond to Middle English /ī/?

This dialectal bind dissolves when we shift our notion of change from something happening to the elements of the system (its sounds, phonemes, morphemes) to modifications taking place in a speaker's competence. His grammar changes from one stage to another. We can compare the grammar of Middle English with the grammar of Modern English and posit certain innovations and restructurings that account for the differences. Grammar rules exist not because of any sort of contrast among them. Middle English [i:] "corresponds" to Modern English [ay] in the sense that the optimal grammar at either stage would derive them from underlying /ī/ and they occur in cognate items, but the only significant comparison is between the grammars and not the sounds or morphs.

Finally, nothing in this paradigm of change requires that simplification in the child generation be preceded by innovation in the parent generation. The two processes are independent in that simplification can occur without prior innovation and parents can add rules to their grammars irrespective of whatever kind of grammar their children are constructing. Children seem to simplify spontaneously. They merely build a grammar based on what they hear. They can have no notion of what the adult grammars look like. A child, in other words, couldn't care less how his parents' grammars got the way they are.

Given the tendency towards optimization, one might well wonder why languages don't end up being maximally simple: three vowels or less, a very few phonological rules, a primitive syntax. We shall briefly enumerate here ways that languages become more complex.

One source of increased complexity is innovation in the adult grammar. This needs no further comment here. It was suggested earlier in this section that children too innovate, perhaps most frequently by adding rules of assimilation. It can be argued that such rules contribute to the over-all simplicity of a grammar since assimilation is a "natural" phenomenon, but by present evaluation procedures grammars are more complex if they contain such rules.

A third source of what might appear to be increased complexity is the collapsing of two or more rules. Any extensive set of phonological rules for a language, such as those listed for English in Chomsky and Halle (1968:238–245), contains rules of considerable complexity, and it is not plausible to assume that such rules entered the language as innovations. Rather, it seems more likely that the innovations were, each taken by itself, relatively simple, but that those innovations affecting the same segment(s) were collapsed into a single rule in later grammars. The resulting rule will then appear complex.

Fourth, certain changes may secondarily complicate other parts of the grammar. Consider a hypothetical language with five underlying vowels /i e a o u/ and the underlying stops /p t k b d g/. Suppose there is a rule lengthening vowels before voiced obstruents, and assume that an innovation

devoicing every /b d g/ is added at the end of the grammar. From underlying /bat/ and /bad/ the surface forms will be [bat] and [ba:t] from earlier [bat] and [ba:d]. The child exposed to these and like forms will hear only length as the distinguishing feature, and we may hypothesize that the child's grammar will have vowel length in underlying forms but nowhere /b d g/. This is a complication of the underlying vowel system, but it also represents the simplest grammar that can be constructed from the output of the adult grammar.

4.4 THE ROLE OF SIMPLIFICATION

In Section 4.2 rule loss was reduced to a special case of simplification. The grammar resulting from rule loss is formally simpler by the number of features in the deleted rule. The output is simpler: more regular, having less allomorphic variation.

Kiparsky (1968b) has proposed that rule reordering too is an instance of simplification, though of a kind different from that discussed so far. Let us examine the case of German rule reordering discussed in Section 3.3. Originally the two pertinent rules, Final Devoicing and Vowel Lengthening, applied in that order, and we would obtain derivations such as the following for *path*:*paths*:

	Underlying Forms:	veg	vegə
(I)	*Final Devoicing:*	vek	.....
	Vowel Lengthening:	.....	ve:gə
	Phonetic Shape:	vek	ve:gə

In the synchronic grammar of German, however, the rules must apply in the opposite order:

	Underlying Forms:	veg	vegə
(II)	*Vowel Lengthening:*	ve:g	ve:gə
	Final Devoicing:	ve:k	.....
	Phonetic Shape:	ve:k	ve:gə

In the original grammar each rule applies once at each step of the derivation. In the grammar containing the rules reordered, Vowel Lengthening has shifted into a position where it applies *twice* at its point of application, and Final Devoicing continues to apply once. That is, the rule of vowel lengthening has moved into a position where it acts on a greater number of forms than previously; the grammar gets the maximum mileage out of this rule, so to speak.

Here, since the grammar contains the same number of rules (and features) whatever the order of the two rules, no notation is available at present to

convert the greater simplicity of one order into greater brevity of the grammar, though such a notation could easily be devised. Kiparsky (1965) has discussed ways in which this might be done; he further has proposed the use of the terms *marked order* and *unmarked order* respectively for orderings (I) and (II). Unmarked order is the optimal order; marked order, the less optimal ordering. The general criterion of directionality in rule reordering may then be stated as: "Marked order tends to be replaced by unmarked order"; or equivalently: "Rules tend to shift into the order which allows their fullest utilization in the grammar" (Kiparsky 1968b).

Note that in the case discussed moving from marked to unmarked order reduces the extent of allomorphic variation in the output. We found this to be true also of rule loss. In marked order (I) 'path' has two allomorphs /vek ~ ve:g/, which differ in vowel length and voice in the final obstruent; in unmarked order (II) the two allomorphs /ve:k ~ ve:g/ differ only in voice value in the terminal obstruent. Of course, this is also true of the large number of words in German with the parallel allomorphy: *Rad* 'wheel', *Bad* 'bath', *lügen* 'to lie', and so on.

In the German case optimal utilization of rules reordered a later rule to an earlier position, where it now applies to a larger number of forms. Section 3.3 discussed another instance of this type: the order of two rules (3.15 and 3.16) in the predecessor language of Old English, Old Saxon, and Old Frisian differed from their order in the other Germanic dialects. It was suggested there that this was a case of reordering, not insertion of a rule into the grammar elsewhere than at the end, though supporting evidence such as relic forms was not available. If we compare these derivations given in Section 3.3 under RULE REORDERING, we see that the chronologically later Rule 3.16 has shifted into an earlier position where it applies to more forms, specifically *þinðanð*. The assumption of reordering here is based primarily on the fact that the direction of the shift is from marked to unmarked order, which is what we find in cases of reordering. (Recall too that all other Germanic languages agreed in having the opposite order.) If the order had shifted the other way—from unmarked to marked, from optimal to less optimal—then we would have a less firm basis for assuming that reordering had occurred. In this case other explanations would merit consideration: either Rule 3.16 was inserted into the grammar elsewhere than at the end of the phonological rules, like Lachmann's Law, or the two rules spread at different rates in a wave-type effect through the Germanic area and reached dialects at different times in the manner discussed in Section 3.3 under RULE REORDERING (cf. King 1968: §3.1).

Cases of rule simplification proper, such as those looked at in Section 3.3 under SIMPLIFICATION, are commonplace in the transmission of language from parent to child. The reason for this is not hard to see; it is as if the child has drawn too much of the right conclusion from the data presented to him. Suppose the parents' grammar has a rule devoicing final fricatives:

$$4.4 \quad \begin{bmatrix} +\text{obstruent} \\ +\text{continuant} \end{bmatrix} \rightarrow [-\text{voice}] \,/\, \underline{\quad} \#$$

In the parents' speech there might then be hypothetical alternations *bif* : *bivo*, *has* : *haza*, *lex* : *lege*. The rule does not apply to stops, so the child would be presented with, for example, *sib* : *sibo*, *wed* : *weda*, *og* : *oge*. Noting alternations such as *bif* : *bivo*, the child correctly intuits that some kind of rule governing the alternation is needed, but his first hypothesis is simpler than that implicit in Rule 4.4. He incorporates into his grammar a rule devoicing all obstruents, whether fricative or stop:

$$4.4' \quad [+\text{obstruent}] \rightarrow [-\text{voice}] \,/\, \underline{\quad} \#$$

In the child's speech we will then have his best efforts to produce *bif* : *bivo*, and so on, but he will also say *sip* : *sibo*, *wet* : *weda*, *ok* : *oge*. This is wrong from the parents' point of view. They don't pronounce things like that. The child has overridden the data and drawn too general a conclusion from them. That children in fact do what we have assumed in this purely hypothetical example is clear. The Ervin (1964) experiment discussed in Section 4.2 showed how children could first learn the correct forms of past tenses (*went*, *ate*, *drank*) and use them, only later to override not only the data but their own previous successful attempts and generalize a rule giving *goed*, *eated*, *drinked*. And this they did even though the weak verbs furnishing the pattern of inflection were infrequent in the speech directed at them.

At this juncture one of two things is possible. The more likely is that the child will eventually reject Rule 4.4′ in favor of Rule 4.4. Presumably the adults in the child's world have only Rule 4.4, and presumably the majority of his playmates have created their grammars with the correct Rule 4.4. Under this pressure the child will complicate his grammar to the extent that he rejects Rule 4.4′ and internalizes Rule 4.4 in the same way that children eventually give up *goed*, *eated*, *drinked*. If, on the other hand, the simplified version, Rule 4.4′, stays in the grammar to adulthood, we have the possibility of a lasting generalization. Certain circumstances favor Rule 4.4′ becoming a permanent, normal part of the language: numerous members of the new generation acquiring Rule 4.4′ in place of Rule 4.4, or final devoicing of all obstruents becoming marked as a prestige item.

It is important to stress that we do not know at present why one simplification takes place rather than the other. Besides Rule 4.4′, why not have a "simplification" of Rule 4.4 such as:

$$4.4'' \quad [+\text{continuant}] \rightarrow [-\text{voice}] \,/\, \underline{\quad} \#$$

(Any continuant is devoiced word-finally.)

From a purely formal point of view this appears to be as legitimate a simplification of Rule 4.4 as Rule 4.4′. Rule 4.4″ devoices all final continuants, e.g. fricatives and vowels, giving *bif* : *bivo̥*, *has* : *hazḁ*, *sip* : *sibḁ*, and so on, where the subcircle denotes voicelessness in vowels. It is possible that rule simplifications like this exist, but it seems rather unlikely; and in the Germanic languages, all of which at one time had Rule 4.4, the putative simplification Rule 4.4′ is unheard of.

The point of this is that there is more to rule simplification than merely deleting features in the structural analysis of a rule. Distinctive features are more than formal signs; they have intrinsic content, and some account of this intrinsic content must be integrated into an adequate phonological theory. In other words, our theory must have some way of stating what formally possible simplifications (like Rule 4.4″) are in fact excluded because they violate principles governing natural languages. Assuming that Rule 4.4″ is in fact an impossibility, an adequate phonological theory would tell us that [+ continuant] can be deleted in the structural analysis of Rule 4.4 to give a simplification, but that [+ obstruent] cannot be. Current phonological theory does not extensively provide us with such constraints, although important steps in this direction have been taken (cf. Chomsky and Halle 1968:400–435).

Whatever the outcome of these efforts to tighten up phonological theory, the historical evidence suggests that rule simplifications obey some kind of hierarchy of features. Features defining major categories like "vowel," "true consonant," "liquid," and "glide" do not seem to be subject to deletion; that is, we do not usually find features like *obstruent* and *consonantal* being deleted in rules. On the other hand, simplifications are permissable by deleting in the structural analysis of rules features presumably low in the hierarchy like *continuant* and *voice* for consonants and *back* or *round* for vowels. Aside from fairly crude observations such as these, it is not possible at present to say much about why one simplification instead of another took place in the development of a language.

Though the bulk of this discussion has dealt with change from generation to generation, some parts of it apply as well to spatial change, in particular the spread of a rule throughout a geographic (or socially defined) area. Borrowed rules are common in the bilingual situation and in the vicinity of a prestige dialect. Thus, speakers from all over the United States who grew up pronouncing final *r*'s sometimes add a rule dropping them if they spend a long enough time in England. We have noted before that adults often add rules to their grammars (innovations), so that there is nothing surprising in the act of rule addition itself. Often, and perhaps generally, such added rules are at first optional, which is reflected in a variation in performance in one's speech. If they become a permanent part of the speaker's grammar, they provide a possible starting point for restructurings in the grammars of subsequent generations. Often such adult rule additions are stylistic in nature: the speaker

uses them to impress his audience or for a raise but normally gets along better without them.

We hypothesize that in borrowing, in general, rules are simplified rather than complicated. That is, a rule is borrowed with the same or greater generality, but not with lessened generality (Harms 1967:172, Bach 1968). Though extensive verification is lacking, this hypothesis has plausibility, and there are some hard data to support it.

Labov (1963) studied the centralization in the first element of the diphthongs /ay/ and /aw/ on Martha's Vineyard and found considerable variation both in the degree of centralization and among social segments of the native population. As Labov (1965:100) concludes, "The centralization of (aw) was part of a more general change which began with the centralization of (ay)." That is, [a] was centralized first in the environment before [y], and as this rule spread over the island of Martha's Vineyard, the environment of centralization was generalized to before [y] and [w]. In terms of the features involved, the environment was simplified from:

$$\_\_ \begin{bmatrix} -\text{vocalic} \\ -\text{consonantal} \\ -\text{back} \end{bmatrix} \quad (\_\_ \text{ y})$$

to:

$$\_\_ \begin{bmatrix} -\text{vocalic} \\ -\text{consonantal} \end{bmatrix} \quad (\_\_ \text{ y, w})$$

Furthermore, there is good evidence (Labov 1963:289–290) that centralization originated in the environment before voiceless obstruents (*right*, *wife*, *night*, *house*, *out*), but in subsequent generations was present in all phonetic environments (*side*, *by*, *I'll*, *now*, *down*). This is again a case of a rule simplification (generalization) during transmission.

A similar phenomenon can be observed in the English of certain Canadian provinces, notably Ontario—particularly in and around Toronto. Some speakers centralize the /a/ in /ay/ and /aw/ only before voiceless obstruents. The environment for centralization has been generalized by many speakers to produce centralization in all occurrences of /ay/ and /aw/. There are Canadians whose dogs go [bʌw wʌw].

It seems a priori unlikely in our theory that a rule would become less general as it is transmitted from one dialect to another. Let us consider again the case of the hypothetical language with Rule 4.4, which devoices only final fricatives. Suppose further that this is a prestige dialect. The speaker of a less prestigious dialect, listening to his betters and wanting to sound more like them, is faced with something like the problem of a child learning the language. He cannot directly tap the grammar of his speech models and pick out Rule 4.4; he can only observe the primary data—their speech—and

formulate a rule to account for what he hears. What he hears are voiceless final fricatives. It seems highly improbable that he would formulate a *less general* rule than 4.4 to add to his grammar, say:

$$4.4''' \quad \begin{bmatrix} +\text{obstruent} \\ +\text{continuant} \\ +\text{anterior} \end{bmatrix} \rightarrow [-\text{voice}] \,/\, \underline{\quad} \#$$

which devoices only labial and dental fricatives, e.g. *v*, *ð*, and *z*. The speaker borrowing the rule will, one may assume, either formulate a rule with equal generality or increased generality (4.4′).

If this hypothesis holds up as more data are amassed, it will give us a useful tool for prying into the linguistic movements of pre-history. If, for example, we know that the living or attested languages A and B share a rule but that this rule is more general in the grammar of A than of B, and if we know that early contact between the two languages existed, then our assumption would be that the rule was transmitted from B (less general) into A (more general) instead of vice-versa.

This assumption runs counter to a widely held view of transmission of rules which holds that rules tend to narrow in generality as they spread farther from the point of origin. The analogy of a stone cast into water insinuates itself here: the ripples are strongest near the center of the disturbance, and they weaken the farther out they go. Our notion of grammar holds no rationale for linguistic behavior like this as a general rule; indeed, the opposite assumption has more inherent credibility within generative grammar. It is as if a stone thrown into water created ripples that grew in strength as they moved away from the center.

A case in point is the High German Consonant Shift, summarized in Section 4.1 with regard to hypercorrection in Low German. This shift—general in South Germany and Switzerland, less general as one travels north, and absent in the native Low German of Northern Germany—has always been regarded as a paradigmatic case of a sound change spreading with decreasing generality (Hockett 1958:480). Our view suggests the opposite direction of transmission: the High German Consonant Shift seems to have begun in the border area between Low and High German (roughly in the area of the "Rhenish Fan"), and to have diffused southward with increased generality. Unfortunately, the dispute cannot be settled because there are too few documents dating from immediately before and after the shift (c. A.D. 500). Cf. Becker (1967:61–64) on this problem.

4.5 A CASE HISTORY: HIGH GERMAN UMLAUT

Let us investigate in detail the progress of a sound change from innovation to restructuring in the development of umlaut—the fronting of back vowels

—in High German from around A.D. 750 to approximately A.D. 1200. The following analysis will provide data for a later inquiry, in Chapter 9, into the relation between scribal practice and phonological representations.

Umlaut was already mentioned in Section 3.3 under SIMPLIFICATION as an example of rule addition and subsequent simplification. All of the surviving Germanic languages show traces of the original process. This is especially true in Standard German, where we still have phonological alternations mirroring the original process: *Kraft* : *kräftig* 'power, powerful', *Lob* : *löblich* 'praise, praiseworthy', *Muße* : *müßig* 'leisure, idle'. Umlaut in the second member of each pair is triggered by the *i* in the suffix. In English, pairs of the type *goose* : *geese*, *foot* : *feet*, *blood* : *bleed* are witnesses to an active umlaut process in earlier English.

It is customary to divide the linguistic history of German into three periods: Old High German (to 1100), Middle High German (1100–1350), New High German (1350 to present). (These dates are only approximate and should serve merely as rough attempts to lend chronological perspective.) In the documents of Old High German only the umlaut of short *a* to *e* is customarily indicated, e.g. *gast* : *gesti* 'guest, guests'; there is no scribal indication of the umlaut of the other back vowels (*ŭ*, *ŏ*, *ā*) during the Old High German period until very late, and then it is sporadic. Thus, corresponding to Modern German *mochte* : *möchte* 'I liked, I would like', we have Old High German *mohta* : *mohti* without umlaut designation in the latter. Nor is orthographic designation of umlaut consistent in the Middle High German period except for short *a*, though we do find in this era increasing scribal inventiveness in orthographic differentiation of the umlaut vowels. In Middle High German we expect to find either *mohte* : *mohte* with no umlaut designation in the latter, or perhaps *mohte* : *môhte*, *mohte* : *möhte*, or some mark to indicate the presence of [ö] in the word for 'I would like'.

The very earliest Old High German documents, those dating from approximately 750 to 800, regularly contain unumlauted short *a*. We reconstruct, therefore, an early stage of Old High German—let us call it pre-Old High German—in which umlaut was not present as a rule in the grammar of the language. At this stage umlaut alternations do not occur at any level, and we have the following derivations:

PRE-OLD HIGH GERMAN

Gloss:	guest	guests	hole	holes	worm	worms
Underlying:	/gast	gasti	lox	loxxir	wurm	wurmi/
Phonetic:	[gast	gasti	lox	loxxir	wurm	wurmi]

We assume that a rule producing umlaut was added to the grammar of pre-Old High German at some time between 750 and 800. We cannot be

certain about the precise form of this rule. The most literal interpretation possible from the written evidence is that it originally affected only short *a*, and then was generalized by Middle High German times to all of the back vowels. But we do not accept this interpretation here; rather, in accord with the consensus of modern and traditional scholarly belief, we assume that *all* back vowels were subject to umlaut even during Old High German, and that the absence of overt umlaut designation in *ŭ̈*, *ŏ̈*, *ā̈* was an orthographic lapse rather than a phonetic one. The rule we posit for the earliest stage of attested umlaut in Old High German has the form:

$$4.5 \quad \begin{bmatrix} \text{V} \\ \langle -\text{long} \rangle \end{bmatrix} \rightarrow \begin{bmatrix} -\text{back} \\ \langle -\text{low} \rangle \end{bmatrix} / \_\_ \ \text{C}_1 \begin{bmatrix} -\text{consonantal} \\ +\text{high} \\ -\text{back} \end{bmatrix}$$

(All vowels are fronted when followed in the next syllable by *ĭ* or *j*; the short vowels thus fronted become nonlow. Thus, *ŭ ŏ ā a*>*ŭ̈ ŏ̈ ā̈ e*. This rule, and the accompanying discussion, ignores complicating details such as "secondary umlaut," the failure of *u* to umlaut in certain dialects, and the presence of umlaut-inhibiting clusters like *hs* and *ht*. Cf. Kiparsky 1965.)

With this innovation came a change in the surface forms of Old High German. A speaker whose grammar did not contain Rule 4.5 said [gasti] 'guests'; one whose did said [gesti]. There has been, however, no restructuring at this point, no change in underlying representations. Whether a speaker said [gasti] or [gesti], the underlying form in the simplest grammar remains /gasti/, and the difference in surface forms arises from application of Rule 4.5. At this stage of history, which we arbitrarily designate Old High German Stage I, typical derivations are as follows:

OLD HIGH GERMAN STAGE I

Gloss:	power	powerful	hole	holes	worm	worms
Underlying:	/kraft	kraftig	lox	loxxir	wurm	wurmi/
Rule 4.5:	.....	kreftig	.....	löxxir	.....	würmi
Phonetic:	[kraft	kreftig	lox	löxxir	wurm	würmi]
Orthographic:	*kraft*	*kreftig*	*loh*	*lohhir*	*wurm*	*wurmi*
Gloss:	deed	deeds	heard	to hear	skin	skins
Underlying:	/tāt	tāti	hōrta	hōrjan	hūt	hūti/
Rule 4.5:	.....	tāti	.....	hōrjan	.....	hūti
Phonetic:	[tāt	tāti	hōrta	hōrjan	hūt	hūti]
Orthographic:	*tāt*	*tāti*	*hōrta*	*hōrian*	*hūt*	*hūti*

Although there has yet been no restructuring in Old High German as we are presenting the development of umlaut, there was restructuring at this stage in an autonomous phonemic account, namely as regards the umlaut of short *a* (Twaddell 1938). Within autonomous phonemics, at the stage of pre-Old High German, /e/ and /a/ had the single allophones [ɛ] and [a] respectively. After umlaut, /a/ had two allophones: [a] normally, [e] under conditions of (primary) umlaut. It is customarily assumed that the umlaut allophone of /a/ ([e]) was phonetically different from the primary allophone of original /e/ ([ɛ]) on the basis of Middle High German rhyme evidence and testimony from the modern dialects, some of which preserve the two *e*'s distinct: [e] from /a/ under umlaut is higher than [ɛ] from original /e/. Autonomous phonemics requires that the [e] resulting from umlaut of /a/ be assigned to the phoneme /e/ because of phonetic similarity. At Old High German Stage I, /e/ had two allophones: [e] under conditions of umlaut, [ɛ] otherwise. /a/ no longer had its umlaut allophone [e] (though it did have a secondary umlaut allophone [ä], which we have omitted from our discussion). In other words, restructuring has taken place in a part of the data under investigation: the underlying autonomous phonemic form of [gesti] 'guests' has changed from /gasti/ in pre-Old High German to /gesti/ in Old High German Stage I, and similarly in all other cases of the umlaut of short *a*. The other umlaut phones [ŭ̈ ŏ̈ ā̈], however, retain their allophonic status as in pre-Old High German. At this stage there has been no restructuring in a generative account, but partial restructuring in an autonomous phonemic account.

Next we shall consider developments subsequent to the stage of Old High German Stage I. One of the umlaut-producing factors begins to disappear in the course of the ninth century: *j*. By the end of the ninth century it is in general lost everywhere except after light syllables ending in *r*. This *j* is written *i* or *e* in the early documents, as we see from a comparison of early and later orthographic forms of words containing *j*:

Early Forms	*Later Forms*	*Gloss*
suntiu	suntu	sin (dative singular)
kennian	kennen	to know
hirteo	hirto	of the shepherds
suntea	sunta	sin (nom. singular)

We formulate then, as an innovation in the grammar of Old High German Stage I, the addition of a rule that deletes *j* in these environments:

$$4.6 \quad \begin{bmatrix} -\text{vocalic} \\ -\text{consonantal} \\ -\text{back} \end{bmatrix} \rightarrow \emptyset \ / \ \_\_ \ \begin{bmatrix} \text{V} \\ -\text{stress} \end{bmatrix}$$

(The glide *j* is deleted when it is followed by an unstressed vowel, e.g. *kennian* [kennjan] > *kennen* [kennen] 'to know'. Of course, *j* remains before stressed vowels as in *iār* [jār] 'year'. We leave out of account in Rule 4.6 the retention of *j* after *r* in light syllables.)

We assume further that Rule 4.6 was added to the end of the grammar after Rule 4.5, the umlaut rule. We now can distinguish a second stage, which we designate Old High German Stage II, dating its inception at approximately 800. This stage differs from Old High German Stage I only in the addition of Rule 4.6; hence derivations such as /kraftig/ > [kreftig] 'powerful' remain unchanged. Only forms containing umlaut under the influence of *j* will have changed surface forms; for example, 'to hear' will now have the derivation /hōrjan/ > [hȱrjan] > [hȱren] (*j* generally raised and fronted following unstressed *a* to *e*). Illustrative derivations of forms from paradigms of 'favor', 'back', and 'sin' follow.

OLD HIGH GERMAN STAGE II

Gloss: 'favor'	*Nom. Sg.*	*Dat. Sg.*	*Nom. Pl.*	*Gen. Pl.*
Underlying:	/anst	ansti	ansti	anstjo/
Rule 4.5:	.....	ensti	ensti	enstjo
Rule 4.6:	.....	.....	.....	ensto
Phonetic:	[anst	ensti	ensti	ensto]
Orthographic:	*anst*	*ensti*	*ensti*	*ensto*
Gloss: 'back'	*Nom. Sg.*	*Dat. Sg.*	*Nom. Pl.*	*Gen. Pl.*
Underlying:	/hrukki	hrukkje	hrukki	hrukkjo/
Rule 4.5:	hrükki	hrükkje	hrükki	hrükkjo
Rule 4.6:	.....	hrükke	.....	hrükko
Phonetic:	[hrükki	hrükke	hrükki	hrükko]
Orthographic:	*hrucki*	*hrucke*	*hrucki*	*hrucko*
Gloss: 'sin'	*Nom. Sg.*	*Dat. Sg.*	*Nom. Pl.*	*Gen. Pl.*
Underlying:	/suntja	suntju	suntjā	suntjōno/
Rule 4.5:	süntja	süntju	süntjā	süntjōno
Rule 4.6:	sünta	süntu	süntā	süntōno
Phonetic:	[sünta	süntu	süntā	süntōno]
Orthographic:	*sunta*	*suntu*	*suntā*	*suntōno*

(Note: the "underlying forms" cited in these derivations are in reality several steps removed from the forms we would take as underlying in a more comprehensive grammar of Old High German. *j* is derived from *i* prevocalically, and the geminate *-kk-* in *hrukki* 'back' is predictable. Thus, the correct underlying form of /xrukkje/ 'back (dat. sg.)', as opposed to the form given here for simplicity of illustration, is /xrukie/.)

At this stage an alteration of considerable magnitude is observed in the speech output vis-à-vis that of Old High German Stage I, yet still no restructuring has taken place within the generative grammar accounting for this speech output. As in the grammar of pre-Old High German, as in the grammar of Old High German Stage I, there has been no change in underlying representations. The optimal grammar at this stage still assumes only ten vowels in underlying forms (*ĭ ĕ ă ŏ ŭ*, but not *ȳ ø̆ ă*), and rules for umlaut and *j*-deletion are present in that order, so that the speech output is full of forms containing umlaut produced by a *j* which since has disappeared. Thus, even in forms such as *hrucko* [hrükko] 'back (gen. pl.)' and *sunta* [sünta] 'sin (nom. sg.)', where no umlaut factor is phonetically manifested, children learning the language at this point would not construct underlying forms containing /ü/. Phonological alternations of various kinds were still present and even plentiful in the data (speech) from which Rules 4.5 and 4.6 could be posited, hence obviating the need for umlaut vowels in underlying forms: [kraft : kreftig] 'power, powerful'; [wurm : würmi] 'worm, worms'; [anst : ensti : ensto] 'favor (nom. sg., dat. sg., gen. pl.)'; [hrükki, hrükke] 'back (nom. sg., dat. sg.)'.

At Old High German Stage II, then, even though to a phonetician it would *sound* strikingly different from that spoken earlier, differences in the grammars are confined to innovations that do not involve changes in underlying phonological representations. From the point of view of autonomous phonemics, on the other hand, there is additional and considerable restructuring at this point, for a number of umlaut vowels attain autonomous phonemic status with the loss of the first *j*. As soon as the first *j* was deleted, the possibility of a contrast between *umlauted* and *not umlauted* exists, and in fact near-minimal pairs can be found in the data: *sunte* : *hunte* [sünte : hunte] 'sin (acc. sg.), dog (dat. sg.)' (cf. Modern German *Sünde* : *Hunde*), *hunte* without umlaut from underlying /hunte/ and *sunte* with umlaut from underlying /suntja/; *māre* : *wāra* [mā̆re : wāra] 'famous (nom. and acc. pl. masc.), truth', *māre* with umlaut from underlying /mārja/, *wāra* without umlaut from underlying /wāra/.

In other words, the front rounded vowels became autonomous phonemic the instant the first *j* in a word like *suntja* 'sin' dropped. And if we adhere to a strong version of the biuniqueness (invariance) condition in phonemic analysis, according to which a phone once assigned to a phoneme must be regarded as a realization of that phoneme each time the phone occurs, we are forced to reassign the umlaut vowels to /ü/ and /ū̆/ respectively in, for example, *wurmi* [würmi] 'worms', *hūti* [hū̆ti] 'skins'. That is, if we accept strong biuniqueness, restructuring occurred already in *all* underlying forms containing umlaut vowels in Old High German Stage II, no matter whether an umlaut-producing factor was still present (like *i*) or not (like *j*).

From this point onward it becomes more difficult to make precise statements about what happened next: the documents tell different, often conflicting stories. Fortunately, however, a scribe of uncommon talent and learning, Notker Labeo, was then at the monastery of St. Gall in Switzerland. His translations into Old High German of such works as Beothius' *De Consolatione Philosophiae* stand today as classics of their kind. He was also a keen phonetician and an inventive scribe: he noted and consistently marked long vowels, which was something none of his predecessors had done with Old High German; he recorded in his translations an external sandhi voicing assimilation present in his speech; and he recorded the reduction of unstressed vowels taking place in Old High German during his lifetime. The latter point interests us most at the moment, for umlaut is contingent upon the status of the unstressed vowels *i* and *ī* since the other factor *j* was no longer present in surface phonetic forms. The most apparent vowel reductions in Notker's writings are unstressed *i u* > *e o*—Notker writes *ubel* 'evil', *fure* 'before', *frido* 'peace', *filo* 'much' instead of the earlier forms *ubil, furi, fridu, filu*. The word *gesti* 'guests' becomes for Notker *geste* with the overt umlaut signal *i* no longer phonetically manifested.

Our question now is: Does this development lead to restructuring? Are we now required to assume underlying front rounded vowels—umlaut vowels—in the underlying phonological representations of Old High German? The answer is again No, and for the following reasons. What was described above as simple unconditioned merger of unstressed *i u* > *e o* was in reality a case initially of partial merger and subsequently a generalization of the merger rule, not a one-step process. The chronology is fairly clear in Notker's writings, and we may summarize it as follows (see Moulton 1961a:29):

> First, unstressed short *i u* > *e o* in checked position: *gestim* > *gestem* 'guest (dat. pl.)', *sibun* > *sibon* 'seven'.
>
> Second, unstressed short *i u* > *e o* in free position: *gesti* > *geste* 'guests', *fridu* > *frido* 'peace'.
>
> Third, unstressed short vowels (now only *e a o*) fell together into [ə].
>
> Fourth, the unstressed long vowels underwent similar reductions, first lowering of *ī* and *ū*, then total merger into [ə]: *zungūn* > *zungōn* > *zungen* 'tongues', *hōhī* > *hōhē* > *hōhe* 'height'.

(Note that in Middle High German, and to some extent in the late period of Old High German we have examined here, *e* in unstressed positions spells [ə] or a reduced vowel similar in quality.)

The mergers of *i* > *e* and *ī* > *ē* are our primary concern here. With the lowering of *i* to *e* we must consider the possibility that restructuring takes place in subsequent generations. The developments among the unstressed

vowels sketched above point to a process of generalization that had the starting point:

$$4.7 \quad \begin{bmatrix} \text{V} \\ -\text{stress} \\ -\text{long} \end{bmatrix} \rightarrow [-\text{high}] \,/\, \_\_ \; C_1 \; \#$$

(The short unstressed vowels *i u* are lowered to *e o* in checked position—that is, when separated from word-boundary by at least one consonantal segment.)

One generalization of Rule 4.7 is caused by an extension of the environment from exclusively checked position to checked and free position:

$$4.7' \quad \begin{bmatrix} \text{V} \\ -\text{stress} \\ -\text{long} \end{bmatrix} \rightarrow [-\text{high}] \,/\, \_\_ \; C_0 \; \#$$

(The short unstressed vowels *i u* are lowered to *e o* in free or checked position—that is, when separated from word-boundary by zero or any number of consonants.)

Rule 4.7′ now is generalized by suppressing the feature [− long] in the structural analysis:

$$4.7'' \quad \begin{bmatrix} \text{V} \\ -\text{stress} \end{bmatrix} \rightarrow [-\text{high}] \,/\, \_\_ \; C_0 \; \#$$

(The unstressed vowels *ī ū* are lowered to *ē ō* in checked and free position.)

This then is the scheme of reduction and generalization that emerges from a consideration of Notker's spellings. The fact that even Notker, consistent as he normally was in his practice, fluctuated in his representations of the unstressed vowels would indicate that Rule 4.7 and its generalizations were originally optional rules in his grammar. This is not at all out of the ordinary. It is quite possible that most innovations occur originally as added optional rules that subsequently become obligatory (Klima 1965:95). Such an assumption accounts for many of the inconsistencies, the exceptions to sound laws of the anti-Neogrammarians, so often found in transition and boundary dialects.

We will for simplicity assume that Notker's grammar contained Rule 4.7 at some point. Of this much we can be sure, though some of the details remain unclear. We then have a stage we shall designate as Old High German Stage III, which we date roughly at 950 to 1050 since the documents show the start of the vowel reduction process at the beginning of the tenth century, and Notker died in 1022. This grammar had derivations such as the following, taken from the paradigm of *gast* 'guest':

OLD HIGH GERMAN STAGE III

Gloss: 'guest'	*Nom. Sg.*	*Nom. Pl.*	*Gen. Pl.*	*Dat. Pl.*
Underlying:	/gast	gasti	gastjo	gastim/
Rule 4.5:	.....	gesti	gestjo	gestim
Rule 4.6:	.....	.....	gesto	.....
Rule 4.7:	.....	.....	.....	gestem
Phonetic:	[gast	gesti	gesto	gestem]
Orthographic:	*gast*	*gesti*	*gesto*	*gestem*

This derivation makes clear the presence of considerable morphophonemic alternation still in the language even after the reduction of unstressed *i* is under way. Furthermore, the umlaut-producing factor is still phonetically present in the paradigms in certain instances, as in *gesti* 'guests (nom. pl.)' above. This would also be true of many other words: *nāmi* [nāmi] 'I, he would take', *nāmīs* [nāmīs] 'you would take', *nāmi* [nāmi] 'you took', all of which have umlaut with *i* or *ī* phonetically present, alternating with *nam* [nam] 'I, he took', *nāmom* [nāmom], 'we took' without umlaut.

From this we conclude that no restructuring of the umlaut vowels had occurred in Notker's grammar nor at this stage in the grammars of his contemporaries in that part of the German-speaking area. Of course, the picture is muddied by the existence of dialects, and here we have made no attempt to present the entire configuration. The point is that the simplest grammar at Stage III has no umlaut vowels in its underlying forms, and the rules 4.5, 4.6, and 4.7 are still present. No restructuring has occurred, though the grammar has become formally more complicated by successive layers of rules since pre-Old High German.

As we have seen, Notker's scribal treatment of the unstressed vowels shows a trend toward generalization of Rule 4.7 which cultimates in Rule 4.7″. Subsequently, all vowels under weak stress merge into schwa—the situation in Middle High German. This reduction is carried through with great consistency in the manuscripts from 1100 on. With the generalization to Rule 4.7″ we reach a fourth stage, Old High German Stage IV, and we date its inception at 1050 or slightly earlier. In the earliest form of this grammar we have derivations similar to those given for Old High German Stage III, but Rule 4.7 is replaced by Rule 4.7″. We then have:

OLD HIGH GERMAN STAGE IV

Gloss: 'guest'	*Nom. Sg.*	*Nom. Pl.*	*Gen. Pl.*	*Dat. Pl.*
Underlying:	/gast	gasti	gastjo	gastim/
Rule 4.5:	.....	gesti	gestjo	gestim
Rule 4.6:	.....	.....	gesto	.....
Rule 4.7″:	.....	geste	.....	gestem
Phonetic:	[gast	geste	gesto	gestem]
Orthographic:	*gast*	*geste*	*gesto*	*gestem*

At this point the great bulk of phonological alternation between unumlauted forms and umlauted forms with an overtly marked umlaut have disappeared from Old High German speech. It is true that some umlaut-triggering *i*'s remained, e.g. *männlich* 'masculine' (base *man* 'man'), *väterlīn* 'daddy' (base *vater* 'father'). That is, a limited number of *i*'s (and *ī*'s) remained when protected by a tertiary stress and under other conditions, and these exist today in Standard German, e.g. *Mann* : *männlich* 'man, masculine', *Hof* : *höfisch* 'court, courtly', but the bulk of umlauting vowels are gone.

Thus, in Modern Standard German—and presumably also in Middle High German and what we have called Old High German Stage IV—there is motivation for regarding some occurrences of umlaut vowels as nonphonemic, that is, derivable by a rule similar to Rule 4.5. Such would be *ä* in *männlich* and *ö* in *höfisch*. There is, however, no compelling motivation for assuming in the optimal grammar of Modern Standard German that *all* umlaut vowels are derivable by phonological rule, in particular those umlaut vowels occurring in monosyllabic, underived words like *schön* 'pretty', *für* 'for', *grün* 'green', and so on. The data from Middle High German point to the same conclusion. (Current phonological theory does not force a clear choice between different treatments of the umlaut vowels in Modern Standard German. Analyses positing no underlying umlaut vowels are possible in the present framework. Here, it is tentatively assumed that some but not all occurrences of umlaut are phonemic.)

We assume, that is to say, that restructuring occurred in the grammars of High German subsequent to Old High German Stage IV and created for Middle High German (roughly from 1100 on) a grammar containing umlaut vowels in underlying forms. The underlying form of [sündə] 'sin' is changed from/ sundja/ to /sünde/, the underlying phonological forms of [gestə] 'guests' and [würmə] 'worms' change from /gasti/ and /wurmi/ to /geste/ and /würme/. Rules 4.6 (*j*-deletion) and 4.7″ (reduction of unstressed vowels) are lost from the grammar, and Rule 4.5 (umlaut) survives as a rule of low "functional load" to account for the forms like *Mann* : *männlich* 'man, masculine'.

Perhaps more than anything else, the foregoing example shows how wide an array of linguistic facts we must appeal to in discussing the diachronic evolution of a language. Literally *everything* in the language is of possible relevance to our analysis—morphophonemic alternation, phonetic changes, morphological processes. All of this affects our decisions at each point in determining what is phonemic, what is predictable by rule, and what the optimal grammar must have been like. We can afford to limit our view to purely phonetic matters, such as phonetic minimal pairs, only at the cost of impoverishing our account of diachronic development.

4.6 SYNCHRONIC GRAMMARS AND HISTORICAL RECAPITULATION

One problem not yet mentioned is the kind of relation that one might suppose to exist between a synchronic grammar and its history. To what extent

does a synchronic grammar recapitulate the historical events that have taken place in the historical grammars of the language? Or, more to the point: Are historical facts relevant to the formulation of the synchronic grammar of a language? We might go even further and put the question as follows: Should a grammar recapitulate the historical development of a language?

Note that questions of this sort, as they usually are asked, arise in synchronic analysis; they have to do with the evaluation of grammars. The underlying sentiment seems to be that a grammar correctly recapitulating the history of a language as well as accounting for the synchronic data is higher valued than one accounting only for the synchronic data. It is obvious, therefore, why we have not concerned ourselves here with questions of this general class: they arise in synchrony, not diachrony. They have to do with constraints that might bear on the evaluation of synchronic grammars, but they do not arise in the discussion of historical change proper. Even so, some discussion of the general question is not totally out of place in a book on historical linguistics, if only because of the frequent misunderstandings of the relation between historical development and the synchronic evaluation of grammars.

Let us begin by asking the question: Does historical evidence decide which of two synchronic grammars is higher valued? The answer is a flat No. Given two grammars G_1 and G_2 that correctly account for the same data, and given that G_1 is simpler than G_2 but that G_2 more nearly recapitulates the historical development, then the simpler grammar G_1 is higher valued than G_2. Given two grammars G_1 and G_2 of equal simplicity, and given that G_2 better reflects the historical development of the language, nevertheless both grammars are equally valued in the evaluation measure. There is no reason to prefer G_2 over G_1 (or vice versa, for that matter). What has been said here about the relevance of historical evidence to synchronic evaluation applies *ceteris paribus* to the evidence of neighboring dialects.

Why evidence of these types is not directly relevant to the evaluation of synchronic grammars should be clear. A grammar is an account of a speaker's intrinsic knowledge of his language, his competence—not his father's competence, not any of his ancestors' competences, not the competence of his neighbor whose dialect is slightly different. To admit historical evidence into the evaluation of synchronic grammars would be to claim that the linguistic competence of one's forebears should play a role in evaluating accounts of one's own competence, and there is no reason in fact or theory to entertain such a curious claim.

Note too that if historical information is allowed to enter into the evaluation of synchronic grammars, the whole question of the relationship between a synchronic grammar and its history ceases to be of any interest. As it stands, it is an interesting, *empirical* question whether a particular grammar recapitulates history; for in cases like reordering, where the grammar does not recapitulate history, we can attempt to determine some general criterion (such

as greater simplicity) that accounts for this. But if historical recapitulation were a criterion in the evaluation procedure, there would not be any point in asking about the relation of a synchronic grammar to earlier grammars of the language. There would be no empirical issue of the slightest interest since, by definition, history has been accounted for in the grammar.

The historical evidence, however, is indirectly relevant to synchronic formulation in a number of interesting and often subtle ways. One of the best proofs of the naturalness of a rule is to show that such a rule occurred as an innovation in a language. This is cogent evidence for a bona fide phonological rule. Likewise, the plausibility of an analysis proposed on strictly synchronic grounds is bolstered if one can demonstrate parallels in the history of the language. A rule CC→C simplifying geminate consonant clusters in English can be motivated on purely synchronic grounds: the rule is needed to produce, for example, correct *dissimilar* [disimilər] from underlying /dis = similær/, compare *dislike*, *distasteful* (see Chomsky and Halle 1968:243). It is comforting to know that this rule was added to the grammar of Early Modern English, for this gives us a minimal guarantee that our analysis is not unnatural. But the sole justification for including this as a rule of contemporary English phonology is synchronic.

Very often knowledge of the history of a language is of considerable help in writing its synchronic grammar. Historical knowledge (as well as knowledge of related dialects) often suggests where to look for phonological alternation, what kind of phonological processes to expect, and so on. Historical knowledge might, for example, suggest the setting up of base forms that are at considerable variance with the surface forms in the language. But the ultimate justification for such a choice rests with simplicity, descriptive adequacy, and the synchronic data. Historical development is useful for gaining insights, but it is not a substitute for the synchronic grammar.

The fact is that synchronic grammars do often enough recapitulate a sizeable part of the history of a language. The Great Vowel Shift rule in the synchronic grammar of English is a case in point. Such recapitulation is not surprising since many rules enter a grammar historically as innovations. If an added rule or some variant of it remains in the grammar for a long time, a synchronic grammar recapitulates history since it must contain that rule to achieve descriptive adequacy. But the rule need not remain in the grammar, nor need its position in the grammar bear true testimony to what happened historically. The rule may be lost. It may be switched out of its original order vis-à-vis another rule. It may be added at a point in the grammar that does not correspond to its chronological order. The rule may be simplified. It may lead to restructuring and then be lost.

All these things may make a synchronic grammar bear not the faintest resemblance to some earlier grammar of the language. When a synchronic grammar does recapitulate history, especially in some subtle and superficially disguised way, we have an interesting but hardly remarkable fact since

historical change is grammar change. When a synchronic grammar fails utterly to reflect history, we have an interesting but equally unremarkable fact. The point is that grammars sometimes tell us a lot about their history, sometimes next to nothing, and sometimes they tell us one thing and history tells us another.

In light of these considerations the proper historical phonology of a language is clearly much more than a set of rules that derive the sounds of, let us say, West Germanic from proto-Indo-European. Even if these rules are made as simple as possible in terms of the distinctive features involved, there is not the slightest reason to suppose that they correspond meaningfully to historical reality. Historical reality includes restructuring, and a simple enumeration of the innovations in a language need not bear any resemblance to what happened historically if the grammar has been restructured. One cannot expect a priori that any innovation will remain in the language as a rule.

A proper historical phonology is the history of the *grammars* of a language, of the competences of successive generations of speakers. The listing of rules converting the sounds of proto-Indo-European into those of West Germanic may be of interest as an exercise in ingenuity and distinctive feature virtuosity, but historical linguistics it is not.

SUPPLEMENTARY READING

Chomsky, Noam, and Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper & Row, Publishers. Chap. 6.

Halle, Morris. 1962. "Phonology in Generative Grammar." *Word* 18.54–72. Reprinted in *The Structure of Language*, eds. Jerry A. Fodor and Jerrold J. Katz (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1964).

Harms, Robert T. 1967. "Split, Shift and Merger in the Permic Vowels." *Ural-Altaische Jahrbücher* 39.161–198.

Kiparsky, Paul. 1965. *Phonological Change*. Unpublished doctoral dissertation, Massachusetts Institute of Technology.

5 SOUND CHANGE AND ANALOGY

Of all the topics of conversation and scholarly research in linguistics that have seen the light of day during the last century or so, surely *sound change* ranks high among those accompanied by nonsense and obfuscation. We are all acquainted with some of the better known examples.

Jakob Grimm supposed that the Germanic Consonant Shift and the High German Consonant Shift were provoked by the impetuous nature of the Germanic tribes—a suggestion that at least one twentieth-century linguistic scholar (Prokosch 1939:55) felt "may fundamentally contain a good deal of truth." Other scholars have discovered considerable merit in the view that both those Consonant Shifts were in part brought about by the increase of the force of aspiration resulting from life in mountainous regions such as the Scandinavian highlands (presumably the *Urheimat* of the Germanic peoples) or the Swiss Alps (where it was assumed that the High German Consonant

Shift originated). Race, physiology, national temperament—all have had their day.

Such examples could be multiplied several fold and discussed at great length, though at no gain for the cause of historical linguistics. What most of these explanations have in common is an almost total fancifulness (since the principles invoked have not been shown to have universal or near-universal validity) and a simplistic putative correlation between cause and effect which must appeal greatly to the hidden child in each of us for these beliefs to have maintained themselves with such tenacity. Serious linguistic scholars have, of course, long since abandoned the more notorious "explanations," but the subject of sound change is still studded with question marks even after a century of hard work. Hardly any statement about the precise character, process, or cause of sound change can be made without challenge from at least some quarter of the linguistic world.

Is sound change necessarily gradual? That is, if [a] changes to [ɔ] in some language, does it take place in a single step [a] > [ɔ], or must it occur over a series of small (perhaps infinitesimal) steps of which the following might be a sample:

$$a > a^{\rightarrow} > a^{\leftarrow} > a > a^{\vee} > ɔ^{\wedge} > ɔ$$

But if the implementation of sound change is gradual, how do we account for such apparent "sudden leaps" as loss, as when initial *#kn-* in *knight* became *#n*; epenthesis, as when usual Old English *brōþor* 'brother' is found written *berōþor*; and metathesis, as when pre-Old English *hros* 'horse' became *hors*? Is sound change completely "regular"; is its occurrence determined by phonetic environment and phonetic environment only? Is it really *sounds* that change, or is it grammar?

Like sound change, *analogy* has long held a prominent place in historical linguistics. The process of analogy however, is less mysterious: when someone (perhaps a child) says *I seed* in place of *I saw*, it seems obvious that he has drawn a false analogy with the regular formations *I kissed, played, dropped*, and so on. Vexed questions of gradualness do not arise. Nevertheless, much is still unclear about analogy, in particular about the conditions under which it takes place. Is a "proportion" a necessary or sufficient condition for analogy? That is, before analogy can take place, must a relation of the form *see* : x = *kiss* : *kissed* (yielding *I seed*) be present? Is there any sense in which analogy is regular?

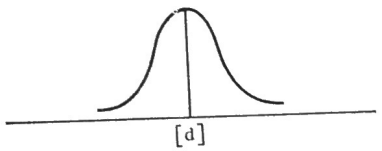
The present chapter will deal with some of the traditional ways of regarding sound change and analogy and will attempt to present a coherent picture within generative grammar.

5.1 THE GRADUALNESS OF SOUND CHANGE

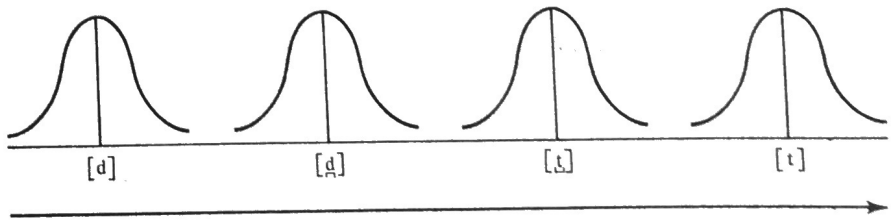
When, for example, Indo-European *b d g* became *p t k* in Germanic, what happened? One view is that gradually during many generations

allophones of /b d g/ came to resemble those of original /p t k/, resulting in a new phonemic series. A change in "habits of articulation" has taken place.

This conception is essentially statistical and may be illustrated by the single change /d/ > /t/. In this view [d], the principal allophone of the phoneme /d/, represents a kind of bull's-eye at which performances of the phoneme /d/ are aimed, much as 50 represents the bull's-eye associated with the number of heads that turn up in a trial of tossing a true coin 100 times. This does not mean that we get 50 heads each time we perform a trial of 100 tosses. Yet in any large number of repetitions of such a trial we expect and find (for a true coin) that the number of heads in each series of 100 tosses tends to cluster around 50, a number we may call (following Hockett 1958:442 and more recently Hockett 1965:194) the "local frequency maximum" associated with the act of tossing a coin 100 times in a row. In this sense the "point" [d] in some abstract articulatory space represents the local frequency maximum, the expected value, associated with performances of the phoneme /d/. The values of performances of /d/ will not necessarily hit the mark exactly, but rather they tend to peak at [d] in accordance with the Law of Large Numbers. We may represent this by the familiar bell-shaped curve:



The sound change /d/ > /t/ consists initially of a random shift of the expected value of /d/ in the direction of [t]. If we assume that [t] lies to the right of [d] in our informal representation, then the initial step in this change would consist of clustering ever so slightly to the right of the previous local frequency maximum of /d/. Since the process is gradual and random, occurring over many years or even generations, no speaker is aware that anything has happened. This process of gradual shifting of local frequency maxima continues, always away from the initial position [d], and the final result is a stable clustering around the value [t]. What we have then is a progression from [d] to [t] over a nondenumerable infinity of local frequency maxima. If from this infinity we select [d̲] (fortis [d]) and [t̲] (lenis [t]) as two representatives, we can represent the process as follows:



Shift originated). Race, physiology, national temperament—all have had their day.

Such examples could be multiplied several fold and discussed at great length, though at no gain for the cause of historical linguistics. What most of these explanations have in common is an almost total fancifulness (since the principles invoked have not been shown to have universal or near-universal validity) and a simplistic putative correlation between cause and effect which must appeal greatly to the hidden child in each of us for these beliefs to have maintained themselves with such tenacity. Serious linguistic scholars have, of course, long since abandoned the more notorious "explanations," but the subject of sound change is still studded with question marks even after a century of hard work. Hardly any statement about the precise character, process, or cause of sound change can be made without challenge from at least some quarter of the linguistic world.

Is sound change necessarily gradual? That is, if [a] changes to [ɔ] in some language, does it take place in a single step [a] > [ɔ], or must it occur over a series of small (perhaps infinitesimal) steps of which the following might be a sample:

$$a > a^{\rightarrow} > a^{\leftarrow} > a > a^{\vee} > ɔ^{\wedge} > ɔ$$

But if the implementation of sound change is gradual, how do we account for such apparent "sudden leaps" as loss, as when initial *#kn-* in *knight* became *#n*; epenthesis, as when usual Old English *brōþor* 'brother' is found written *berōþor*; and metathesis, as when pre-Old English *hros* 'horse' became *hors*? Is sound change completely "regular"; is its occurrence determined by phonetic environment and phonetic environment only? Is it really *sounds* that change, or is it grammar?

Like sound change, *analogy* has long held a prominent place in historical linguistics. The process of analogy however, is less mysterious: when someone (perhaps a child) says *I seed* in place of *I saw*, it seems obvious that he has drawn a false analogy with the regular formations *I kissed, played, dropped*, and so on. Vexed questions of gradualness do not arise. Nevertheless, much is still unclear about analogy, in particular about the conditions under which it takes place. Is a "proportion" a necessary or sufficient condition for analogy? That is, before analogy can take place, must a relation of the form *see* : x = *kiss* : *kissed* (yielding *I seed*) be present? Is there any sense in which analogy is regular?

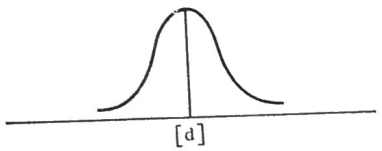
The present chapter will deal with some of the traditional ways of regarding sound change and analogy and will attempt to present a coherent picture within generative grammar.

5.1 THE GRADUALNESS OF SOUND CHANGE

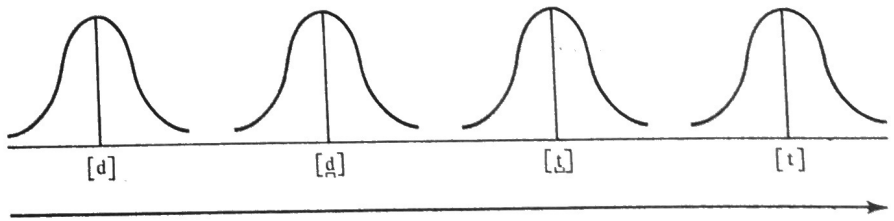
When, for example, Indo-European *b d g* became *p t k* in Germanic, what happened? One view is that gradually during many generations

allophones of /b d g/ came to resemble those of original /p t k/, resulting in a new phonemic series. A change in "habits of articulation" has taken place.

This conception is essentially statistical and may be illustrated by the single change /d/ > /t/. In this view [d], the principal allophone of the phoneme /d/, represents a kind of bull's-eye at which performances of the phoneme /d/ are aimed, much as 50 represents the bull's-eye associated with the number of heads that turn up in a trial of tossing a true coin 100 times. This does not mean that we get 50 heads each time we perform a trial of 100 tosses. Yet in any large number of repetitions of such a trial we expect and find (for a true coin) that the number of heads in each series of 100 tosses tends to cluster around 50, a number we may call (following Hockett 1958:442 and more recently Hockett 1965:194) the "local frequency maximum" associated with the act of tossing a coin 100 times in a row. In this sense the "point" [d] in some abstract articulatory space represents the local frequency maximum, the expected value, associated with performances of the phoneme /d/. The values of performances of /d/ will not necessarily hit the mark exactly, but rather they tend to peak at [d] in accordance with the Law of Large Numbers. We may represent this by the familiar bell-shaped curve:



The sound change /d/ > /t/ consists initially of a random shift of the expected value of /d/ in the direction of [t]. If we assume that [t] lies to the right of [d] in our informal representation, then the initial step in this change would consist of clustering ever so slightly to the right of the previous local frequency maximum of /d/. Since the process is gradual and random, occurring over many years or even generations, no speaker is aware that anything has happened. This process of gradual shifting of local frequency maxima continues, always away from the initial position [d], and the final result is a stable clustering around the value [t]. What we have then is a progression from [d] to [t] over a nondenumerable infinity of local frequency maxima. If from this infinity we select [d̲] (fortis [d]) and [t̲] (lenis [t]) as two representatives, we can represent the process as follows:



It is rather as if as one tossed a coin it began to wear on one side so that the probability of getting heads decreased gradually from 0.50 through 0.47 and 0.42 to 0.40, whereupon the coin quit changing its weight and shape and began to turn up heads consistently 40% of the time.

In this way we have arrived at a new phoneme /t/ with the principal allophone [t]. As before with /d/, the random nature of the articulatory process will lead to a normal curve (bell-shaped curve) distribution with its maximum at [t]. Sound change is still going on in the sense that not every performance of /t/ hits the bull's-eye [t] exactly, but the expected value remains relatively constant, at least until some new trend sets in and carries the most likely value of /t/ away from [t] and towards a new frequency maximum, say [t^h], [t^s], or [θ].

Several implicit assumptions in this picture of sound change should be emphasized. First, sound change is gradual and imperceptible to a single speaker or perhaps to all the speakers of a single generation. Second, sound change is constantly in progress since performances of a given phoneme differ each time. Third, the speaker's competence—competence in the technical sense, his implicit knowledge of the language—is irrelevant to the process. Sound change is not change in competence but change in performance brought about by external factors that affect and alter renditions of a particular sound type: the amount of moisture in the vocal passages of the speaker, his muscular tone, whether he is drunk or not, and so on (Hockett 1958:443).

Phonological change as it is conceived of in generative grammar differs radically from this. First and foremost, as has been emphasized throughout this book, change is change in competence reflected by alterations in the grammar. The role of performance remains the same, causing the same kinds of fluctuations after the change in competence as before. To be more explicit, in this uncomplicated case of innovation we assume that a rule $d > t$ has been added to the speaker's grammar. Where he previously said *d* he now says *t*, and we register this fact in our account by the addition of a rule—a change in competence. Before the innovation, realizations of /d/ doubtless did fluctuate in various ways because of the presence of moisture in the vocal apparatus, the speaker's alertness, and so on; but precisely the same performance factors are active after the innovation as before. Their relation to the change in competence is one of complete neutrality—they neither caused it, contributed to it, abetted it, nor slowed it down. The performance factors simply cause the random fluctuations that always take place in articulating sounds and account for the often heard statement that "each speech act is a unique event: no two pronunciations of the same sound are ever the same, even when pronounced by the same speaker."

Second, nothing in generative grammar requires or supports the assumption that the change was necessarily gradual. We simply assume that the rule changing [d] to [t] was added to the speaker's grammar; this changed [d] to [t]. Nothing in fact or in the theory of generative phonology suggests that this change, or any other, had to take place as a series of rule additions of the type: [d] → [+ 3 voice], [d] → [+ 2 voice], [d] → [+ 1 voice], and so on.

The statistical model of gradual sound change outlined earlier assumed that change was infinitesimal over a continuum. This strong version of gradualness can be weakened by dropping the requirement of infinitesimal change; rather we posit "small" changes, where "small" is understood as meaning "within the limits set by a given phonetic alphabet and its associated diacritic marks."

Evidence will now be presented that the gradualness assumption, in either of its formulations, is untenable as a necessary condition on sound change. Furthermore, it will be argued that the term "sound change" is an improper concept for the phenomenon to which that designation is customarily applied. Sounds don't change; grammars do. These are substantive issues, not terminological ones, and we must look to the empirical evidence for confirmation.

A number of linguistic facts support the claims that gradualness is not a necessary condition for sound change and that grammars, not sounds, change. First, there is the indisputable existence of cases such as loss, metathesis, and epenthesis in which any kind of gradual process strains the imaginative faculties as well as the set of distinctive features that one assumes to be universal. Let us consider loss. Loss of segments is an almost commonplace kind of historical development: Greek lost its final stops, Germanic lost word-final consonants and vowels under certain conditions. If, say, *t* is lost word-finally, we account for this simply by assuming that a rule:

$$t \rightarrow \emptyset \; / \; \_\_\_ \; \#$$

was added to the grammar of one or more speakers as an innovation. Perhaps the rule spread within the speakers of a single generation as a fashionable way of pronouncing things; perhaps subsequent generations of speakers restructured their grammars so that no instances of word-final *t* were derivable. The eventual result is that no one pronounces word-final *t*'s. It is true that one can postulate some sort of undeniably gradual process, for example:

$$t > \underline{t} > \theta > \underline{\theta} > \emptyset$$

where ⌣ denotes laxness of articulation. But in cases of loss like this there is never unambiguous evidence in the form of scribal testimony or dialectal variations that would clinch the argument for a gradual process. What we find is that a consonant was in full force in one stage of the language and gone later.

One might at this point argue that the testimony of historical linguistics is suspect in regard to the gradualness of sound change since no one was around to hear what people were saying and since no scribe would be apt to render

a faithful phonetic record of the intermediate stages in sound changes. But there are cases of loss observable in our own day in which no progression of sounds intermediate between the end points can be observed. Some speakers of American English pronounce their final *r*'s, some don't. There is no indication of a gradient of sounds along the progression from [r] to zero or schwa, that is, from [fa:r]>[fa:] or [fa:ə̯] 'far'. To cite here "*r*-colored vowels" as evidence of gradualness is to beg the question. Even if we assume (and there is no compelling reason to do so) that the loss of *r* in many varieties of English took place through a stage containing an *r*-colored vowel, e.g. *r*-colored schwa [ɚ], then we have the progression:

$$r > ɚ > ə$$

and now the proponent of gradualness must find a progression of sounds intermediate between [r] and [ɚ] and between [ɚ] and [ə].

Another instance in American English of supposedly gradual sound change is the so-called intervocalic "voiced *t*" in *water, latter, sitting, batted.* The phonetic facts concerning this sound summarized by Heffner (1960:129–130) indicate a situation as follows. Some speakers of English have voiceless [t]; some Americans have voiced [d] not distinguishable from the pronunciation of *d* in *ladder, shudder*; and some have "voiced *t*" which differs from voiced *d* in minor phonetic details of tenseness and/or duration of hold. This evidence supports the explanation that some speakers of American English realize intervocalic /t/ as [t̬], whereas other speakers realize it as [d]. Still other speakers of English, those who do not voice their intervocalic *t*'s at all, realize /t/ intervocalically as [t] or [tʰ].

Data like these are not prima facie evidence for the gradualness of sound change; they merely support the claim denied by no one that speech variation exists in a natural language. To use this in support of gradualness one would have to demonstrate that speakers with [d] earlier had [t̬], and that speakers with [t̬] earlier had [t] or [tʰ]. It would even suffice to show that one generation had [t] or [tʰ], a later generation [t̬], and a still later generation [d]. But no one has tried to show this, and informal observations of the phenomenon of intervocalic /t/ do not suggest a regular gradient by age of the sort envisioned above.

A second, similar case from English involves the pronunciation of initial *wh* in *when, whether, why, what.* This was discussed in Section 4.3 as an instance of innovation and subsequent restructuring, in which a rule deleting *h* before *w* was added to the grammar:

$$5.1 \quad \begin{bmatrix} -\text{vocalic} \\ -\text{consonantal} \end{bmatrix} \rightarrow \emptyset \; / \; \_\_ \; \begin{bmatrix} -\text{vocalic} \\ -\text{consonantal} \\ +\text{voice} \\ +\text{back} \end{bmatrix}$$

As Heffner (1960:161–162) points out, there are quite a large number of variants of initial *wh*, so that the change expressed by Rule 5.1 is very much an oversimplification. Heffner gives the following variants, where the sub-dot denotes voicelessness in the segment and the sub-omega labialization: [w], [w̥], [hw̥], [h̥ω̥], [x̥ω̥], [hw]. One might regard these as a "sizeable portion" of a large number of values between [hw] and [w]. But again this begs the question of the progression of intermediate values lying between *these* putative intermediate values: can it be shown that one generation had [hw̥], the next [w̥], the next [w]?

Minor variation in the performance or realization of sounds is simply an aspect of language; there is nothing which requires us to regard such variation as *causing* sound change. The amount of aspiration on initial /p t k/ in English varies from much to very little, both within the speech of a single speaker and between speakers. This has no doubt been true for many centuries; but no sound change has affected initial /p t k/ during this time, nor do current prospects for such change seem auspicious. And similar statements could be made about the majority of sounds in any given language at any particular time. Infinitesimal variations in the realization of sounds are present, always have been, always will be; but it requires empirical confirmation in the form of a generational gradient to show that they are the *mechanism* by which sound change occurs.

Linguistic literature is full of cases of radical change in which no intermediate stages are alluded to. Hermann (1931:15, 33) cites one case in Yaghnobi where children have [i] in place of their parents' [e] and a second case in Frisian dialects where some speakers have [l] and others [d]. In neither case is any intermediate value indicated. Gabelentz (1901:193) cites a study of the Samoan languages in which the spread of a sound change $t > k$ is discussed. In less than forty years, hardly more than a single generation, the sound change had been carried out, and there is no evidence for an intermediate stage or series of intermediate stages.

Even if cases like these were postulated as gradual—if, in other words, the failure to observe the postulated gradient of intermediate sounds were ascribed to insensitivity in the human ear or to crudeness in phonetic and acoustic instruments—there are still certain categories of sound change in which gradualness is even more radically counter-intuitive and unreasonable. Such a category is metathesis (the interchange of two segments). Metathesis is not uncommon as a historical change, and synchronically it is found as a rule in the grammars of currently spoken languages (cf. Chomsky and Halle 1968:360–362). Modern English *third*, Old English *ðirda* comes from earlier *ðridda* via metathesis, as does Modern English *horse* from earlier *hros* (cf. Old High German *hros*, Old Icelandic *hross*). What kind of gradual change by allophones can be imagined here or in any case of metathesis?

What kind of gradualness is reasonably possible in epenthesis? If null becomes the vowel [e] in some environment, is it not simpler to assume the

addition of a rule $\emptyset \rightarrow$ [e] rather than some hypothetical and completely implausible gradient of sounds between nothing and [e]?

Similarly, in present-day instances of assimilation there is no evidence of gradualness. When, for example, in Spanish $n > m$ via assimilation as in *San Pedro* [sampeðro], there is no gradual realignment of allophones. Are we to assume that historical sound change was fundamentally different in nature?

A further kind of sound change is even more fundamentally incompatible with the gradualness assumption: the changes effected by the so-called "exchange rules" or "alpha-switching rules" of phonology. One of the best known examples is the Great Vowel Shift in English, briefly discussed in Section 4.3. (On the Great Vowel Shift and exchange rules in general see Chomsky and Halle 1968:254–259. See also Wang 1968.) In this set of changes affecting the tense vowels of fifteenth-century English, high and mid vowels "exchanged" places: $\bar{\imath}$ and $\bar{u}$ were lowered to $\bar{e}$ and $\bar{o}$, original $\bar{e}$ and $\bar{o}$ were raised to $\bar{\imath}$ and $\bar{u}$. The rule for these changes may be stated as:

$$5.2 \quad \begin{bmatrix} \alpha \text{high} \\ -\text{low} \end{bmatrix} \rightarrow [-\alpha \text{high}] \; / \; \begin{bmatrix} \underline{\quad} \\ +\text{tense} \\ +\text{stress} \end{bmatrix}$$

In general, exchange rules have the schematic form:

$$[\alpha F] \rightarrow [-\alpha F] \; / \; \ldots$$

where F denotes a feature whose value is switched from + to − and from − to +. In Rule 5.2 when α is + the rule applies to high vowels, making them [− high]; when α is − mid vowels are affected, becoming [+ high].

Wang (1967:102) has presented several cases in Chinese dialects of exchange rules that involve the switching of tones. One such rule carries out the following changes: high tone becomes low tone, low tone becomes high tone, and mid tone is left unchanged. The rule which makes these changes is an innovation as follows, where the features used are all binary features of tone:

$$5.3 \quad [\alpha \text{ HIGH}] \rightarrow [-\alpha \text{ HIGH}] \; / \; \begin{bmatrix} \underline{\quad} \\ -\text{CENTRAL} \end{bmatrix} \begin{bmatrix} +\text{HIGH} \\ +\text{FALLING} \end{bmatrix}$$

Examples of exchange rules, both synchronic and diachronic, could be given from a variety of other languages. Their existence seems not subject to dispute; and there is no way in generative phonology to exclude them as lawful innovations in the grammar of a language. But if this is so, how can we possibly account for them in a theory of change that requires sound change to be gradual, incremental, and infinitesimal? How could high and low tones switch in a language containing a mid tone without disastrous

confusion in the process? The most obvious answer is that such changes are just not gradual: they are phonetic leaps, as it were, and no merger takes place because the sounds or tones being switched never pass through the same point.

It is always possible to assume that two sounds switching places in an exchange rule take different paths, thereby avoiding merger. In explaining how $\bar{\imath}$ and $\bar{e}$ did not merge in the Great Vowel Shift, one might postulate intermediate changes as follows, where the numbers denote chronological sequence of change:

$$\begin{array}{ccc} & 4 & \\ \bar{\imath} & \leftarrow & \bar{u} \\ 2 \downarrow & & \uparrow 3 \\ \bar{e} & \rightarrow & \bar{o} \\ & 1 & \end{array}$$

This way of explaining sound interchanges is always available; but in the absence of confirming data, one must view such explanations as unacceptable. An "alternate route" explanation generally is advanced merely to salvage the gradualness assumption and not because of hard evidence in the form of scribal records or dialect variation (see the remarks on the Great Vowel Shift quoted in Chomsky and Halle 1968:255). This is a high price to pay for retaining an assumption whose appeal is not irresistible to begin with.

That sound change is neither necessarily nor in general gradual is, of course, not a new view, nor is it the exclusive property of generative grammarians. Sommerfelt (1923) wrote in support of abrupt sound change. Hoenigswald (1960:73) suggests that the notion of gradual sound change is a remnant from pre-phonemic days (see also Hoenigswald 1964). Jakobson (1931:249) wrote of the "abrupt character of phonological changes."

The evidence so far presented argues against gradualness as a necessary condition for sound change. In a deeper sense, however, this is a subordinate question; the real question is whether sound change in the traditional sense is a proper concept at all: is it the *sounds* that change, or is it something else? Generative grammar maintains that it is not sounds or phones that change but grammar—a speaker's competence (Postal 1968:269–307). Alteration in competence is reflected by alteration in performance, but not the other way around. Sound changes result from changes in competence, in the internalized system of rules for linguistic behavior. Such changes are of various kinds; rule additions and losses, reordering, simplification. These changes are not caused by tiny variations in performance that somehow seep osmotically up into competence and change it. To use a concrete example,

consider the problem of *wh* in English. Whatever segment underlies this sequence (Chomsky and Halle 1968:223–224 propose /x^{w}/), speakers of English have rules producing a phonological surface-level sequence [hw]. This sequence is then realized by low-level, possibly n-ary rules in the phonological component. Speakers who habitually say [hʷ] have a late rule devoicing the [w]. Speakers who alternate freely between [hʷ] and [ʷ] have an optional phonetic rule deleting the [h].

In other words, variations like these are due to minor differences in late rules in the phonological component of the grammar. We do not at present know very much in detail about performance factors affecting phonetic output, but it is probable that variations such as [hw]~[h̬] for *wh* are best accounted for not by rules in a particular grammar but within a universal theory of performance. That is, it should turn out that certain kinds of minor phonetic variation are universal or near-universal. If it is true that the sequence [hw] in language can be indifferently realized as [ʷ] or [h̬], then this variation is not a part of the grammar at all but is accounted for in the theory of performance associated with the grammar.

Note further that the traditional theory of *sound* change (as opposed to *grammar* change) is indirectly disconfirmed by the existence of nonchronological rule additions like Lachmann's Law in Latin (Section 3.3). The change of *a* to *ā* yielding *āctum* cannot be stated by rules changing the phonetic surface-level sound [a]; Lachmann's Law required the representation [agtum] which never occurs phonetically in Latin. Sound change (i.e. "phone" change) fails to account for this, but there is no problem here for grammar change.

In view of these considerations it would be more preferable to replace the traditional term "sound change" by something more appropriate. The designation "phonological change," which conveys a sense of something more abstract than change by phones, is perhaps the most suitable candidate.

The chief underlying reasons for the assumption of gradual sound change seem to be (a) the sentiment that communication would break down if change were not gradual, and (b) the observation that people are not normally aware of change. There is no empirical confirmation of (a). Languages have an amazing ability to undergo change without impairing communication. Mergers occur, yet communication goes on. Some languages have different varieties of male and female speech (Sapir 1929), yet the sexes communicate. Speakers from the Southeast United States who do not differentiate *i* and *e* before nasals go north and west, but communication does not break down.

Assumption (b), that ongoing phonological change is unnoticed, is doubtless frequently true but, as Chomsky and Halle (1968:250) point out, only because speakers are in general unaware of the contents of their grammar, whether change has occurred or not. The average speaker of a language pays very little attention to pronunciation in the normal course of things; he may notice he's talking to a foreigner or some dialect speaker, but normal

speech variation goes mostly unnoticed. Someone who has added a rule to his grammar doesn't feel any different, and he is usually not even aware that he sounds any different. Similarly, a child who has simplified his grammar does not feel he has done something different or naughty.

In short, there is no logical reason why phonological change must be gradual nor empirical confirmation that it is. Subject to counter-evidence yet to be produced, we reject gradualness as a necessary condition for the implementation of any phonological change.

Parenthetically, let it be noted that linguistic change other than phonological is clearly *not* gradual by any stretch of the imagination. If an adult learns to use *whom* in place of *who* in the right places, how could this be anything but sudden and abrupt? When a child says *foots* instead of *feet*, what is gradual about it? In semantic change, as when *in* takes on the additional meaning "fashionable" or "what everyone else is doing," we are not dealing with gradual change. Likewise, it is hard to imagine accentual changes—like shifting place of accent—as occurring gradually.

Some of the puzzling phonological changes in history lose their apparent mystery once we abandon the unsupported notion of gradual phone change. One of these is the change of Latin *ct* [kt] > Rumanian *pt*, e.g. Latin *octo* > Rumanian *opt* 'eight'. This is part of a more general process whereby in addition the velar nasal [ŋ] becomes [m] before dental stops (including here nasals), e.g. Latin *lignum* [liŋnum] > Rumanian *limn* 'wood' (Kiparsky 1965). The problem is not complicated; a rule was added to the grammar of Rumanian of the form:

$$5.4 \quad [-\text{continuant}] \rightarrow [+\text{anterior}] \; / \; \_\_ \; \begin{bmatrix} -\text{continuant} \\ +\text{coronal} \end{bmatrix}$$

(Stops, including *k* and the nasal *ŋ*, become labial before dental stops.)

This change in competence is reflected by a change in performance: the speaker stops saying *kt* and starts saying *pt*. The rule may first have been optional, so that the speaker was inconsistent in its application and fluctuated between *kt* and *pt*, but the Rumanian evidence shows that it eventually became obligatory. The result is a change of *kt* > *pt*, not different in substance from any other less striking change such as Indo-European *b d g* > Germanic *p t k*. This phonological change in Rumanian is puzzling only within a theory of change that requires gradualness, where it is hard to see how *k* could have become *p* or *ŋ* become *m* without being confused with *t* and *n* along the way. One can imagine an "alternate route" gradual process which would avoid this, e.g. *k* > *kʷ* (with labialization) > *p*, or one may attribute this change to borrowing of various kinds (Naert 1941). But such assumptions result from the assumption of necessary gradualness, and once we have discarded this assumption, they become gratuitous.

Another puzzling change of this sort, which is really no more puzzling

than any other phonological change once gradualness is dropped, has already been alluded to in Section 4.5: the umlaut of Old High German *a*. The umlaut of *a* is assumed to have been a higher mid front vowel [ɐ], different from inherited *e* which is assumed to have been [ɛ]. The poets of Middle High German, as is copiously attested, were careful not to rhyme the two, and we even have a fair number of minimal pairs from Middle High German: *wegen* [wɛgən] 'to weigh': *wegen* [wegən] 'to move'; *her* [hɛr] 'sir': *her* [her] 'army'. That umlaut-*e* differed in tongue height from inherited *e* is assumed on the basis of testimony of modern German dialects that, unlike Standard German, maintain the two distinct. If the change of [a] to [ɐ] under umlaut conditions was gradual, necessarily at some point in this development the two *e*'s were phonetically indistinguishable and would have merged. One could, of course, propose a route such as [a] > [ɔ̈] > [ɐ], which would prevent a "collision" between old [ɛ] and the umlaut of [a], but neither the written documents nor the modern dialects lend credence to this theory. Instead, it has been tacitly assumed that the path from [a] to [ɐ] via umlaut led over [ɛ]. How could this have happened without leading to merger of the two sounds? Solutions have been offered (Fourquet 1952). The point is, however, that a problem exists here only if gradualness is assumed. If we do not require gradualness and if we regard this change as the addition of a rule changing [a] to [ɐ] in the umlaut environment (Rule 4.5), then this change is no different from any other such as metathesis, epenthesis, or simply $p > p^h$.

One major point to be made here is that gradualness leads to a fracturing of the picture of diachronic phonological change without conferring any corresponding benefits. If we take gradualness as a necessary condition for at least *some* phonological changes, say simple shifts among single vowels and consonants, then we are forced to establish at least two categories of change: (1) cases of gradual change $x > y$, and (2) cases of nongradual change $x > y$ such as metathesis, epenthesis, and loss. To the latter category we must then assign a mechanism of change that differs in substance from that of the former category, for example, borrowing or perhaps a different kind of analogy.

Considerations such as these have led scholars to regard some sound changes at least as special cases of borrowing (Hoenigswald 1960:55) and to regard even borrowing itself as a special case of analogy (Chafe 1961:117). Within generative grammar there is no formal distinction between borrowing and spontaneous innovation in a single dialect or in the idiolect of a single speaker. In either instance we would be faced with a change in competence—a rule added to the grammar—and any effort to assign the description "borrowed" or "spontaneous" to the rule would not be relevant to the change itself and its subsequent ramifications. The question of whether an innovation was borrowed or sprang up independently is, of course, not devoid of interest, especially in establishing a genetic relationship on the basis of shared features. But there is no reason in generative grammar to distinguish between changes

that are regarded as gradual (nonborrowed innovations) and those that are clearly not gradual (borrowed innovations).

Rejecting the gradualness assumption does not force one to exclude a priori the existence of intermediate steps. If, for example, *s* in an early stage of a language is represented by zero in a later stage, we are not compelled to assume the addition of a rule $s \rightarrow \emptyset$. The sequence $s \rightarrow h \rightarrow \emptyset$ is more likely: $s \rightarrow h$ is a natural change, as is $h \rightarrow \emptyset$. That is, in certain changes increments may be more expected, more natural than great leaps. But intermediate steps are not a necessary condition, and positing *h* between *s* and zero is not the same as positing an age gradient of sounds between *s* and zero.

It should perhaps be stated expressly that in denying the gradualness of phonological change we do not deny the possibility that a phonological change *spreads* gradually throughout a speech community. These are two totally different things, and our stance with one of these questions in no way commits us to a position with the other. In fact, all the evidence agrees that the *spread* of a change is gradual to a greater or lesser extent. Isoglosses usually move about gradually over periods of time: a favored pronunciation spreads out from a prestige focal point, an archaic feature of pronunciation recedes under pressure from increased communication, from schools, from radio and television. Generally in these cases a phonological change takes place—typically a rule is added—and then the rule is gradually acquired in the grammars of an ever-increasing number of contiguous speakers. This process is assuredly gradual, but it has nothing to do with the question of whether phonological change originates in a constant, gradual, imperceptible shifting of allophones.

To make the difference perfectly clear, let us consider an example briefly discussed in Section 4.4: the spread of centralization in the diphthongs /ay/ and /aw/ on Martha's Vineyard (Labov 1963). The facts are clear. Centralization has come to mark its possessor as "belonging" on the island, as being a bona fide Martha's Vineyarder in contrast to the many tourists and summer visitors from the mainland. Careful linguistic interviews from 1933 show some centralization in occurrences of /ay/ and virtually no centralization in occurrences of /aw/, but in 1963 centralization has spread in such a way that the oldest speakers (over 75) have the least amount of centralization, those from 61 to 75 have more, and centralization increases down to the speakers between 31 and 45. "Amount of centralization" is here a measure of two factors: (1) the degree to which the subject centralizes, i.e. whether he says *right* [rəˇɨt], [raˆɨt], or [raɨt], and (2) the frequency with which he centralizes at all, i.e. the number of times in the speech sample that he pronounces *right* with centralization of any degree as opposed to *right* with no centralization [raɨt]. The data presented in Labov (1963) bring out several relevant factors: centralization is desirable; there is an age gradient in amount of centralization; certain phonetic environments (before *t*, *s*) favor centralization over others (before *m*, *n*); the phonological rule expressing centralization is

optional with some percentage of the speakers (they can but need not centralize, depending on the situation); even in repeated performances of the same word different degrees of centralization are audible (this seems to depend on stress). (Labov's data show two degrees of centralization between [a] and [ə].)

Note especially that "amount of centralization" has two components. The age gradient showing that amount of centralization varies inversely with age thus does not constitute evidence for a gradual shift in the "habit of articulating" /ay/ and /aw/ through generations. What it does demonstrate is that most older speakers do not centralize at all in pronouncing most instances of /ay/ and /aw/, whereas younger speakers do. Significantly, individual speakers fluctuate between [aɪ], [a˄ɪ], and [ə˘ɪ] in their own speech performance (Labov 1963:287–289). This is evidence for fluctuation in performance; it is not evidence for a gradual shift over time in the habit of articulating /ay/ and /aw/.

We may interpret this as a case of the spread of a rule in a more general and slightly altered form. We assume that some speakers, presumably older inhabitants whose identity with Martha's Vineyard—its pace of living, its ideals, and so on—was total and unquestioned, had a rule of centralization in their grammars:

$$5.5 \quad \begin{bmatrix} \text{V} \\ -\text{round} \\ +\text{back} \end{bmatrix} \rightarrow [-1\ \text{low}] \,/\, \underline{\quad} \begin{bmatrix} -\text{vocalic} \\ -\text{consonantal} \\ -\text{back} \end{bmatrix} \begin{bmatrix} +\text{obstruent} \\ -\text{voice} \end{bmatrix}$$

(The [a] in the diphthong [aɪ] is slightly centralized when followed by a voiceless obstruent. This is, of course, an idealization in that some speakers doubtless had a slightly less general rule, perhaps centralizing only before *t* and *s* or before *t*, *s*, *p*, and *f*; some speakers a more general rule centralizing [a] in both [aɪ] and [aʊ]; and some a rule with greater centralization [−2 low] in general or depending on environment. Note too that this is a late phonetic rule in which we may freely use n-ary variables.)

On Martha's Vineyard since 1933 Rule 5.5 has been borrowed in all sorts of varying forms. It is a sign of "belonging" to have Rule 5.5 in one's grammar; it is a desirable acquisition. In the grammar of some, and possibly most, it is an optional rule whose application depends on factors such as stress: Labov (1963:290) reports a fisherman using the word *knife* twice in the space of a few seconds, once [naɪf] and once [na˄ɪf]. The environment of centralization has been generalized to include [aʊ], and it is not necessary that the diphthong be followed by a voiceless obstruent. In some grammars the rule has been borrowed with a difference in the structural change; there is greater centralization, which is reflected in the rule by the specification [−2 low], [−3 low], or whatever degree of minus lowness a particular speaker uses in a particular environment.

In short we may assume this to be a rather typical paradigm of the spread of a phonological change. An innovation occurs or is present from some source and becomes desirable. The rule expressing the innovation spreads, characteristically in a more general form. At first the added rule is optional but it becomes obligatory if the innovation has "sticking power." Minor variations in the structural change of the rule may occur as it is borrowed among speakers. After a while we have an accomplished fact—a sound change.

But note that nothing in all this forces us to assume that the change was gradual in the sense discussed earlier. Its *spread* is in general gradual, not sudden, and to describe it we should properly reach for statistical tools showing the percentage of speakers in an age group that have the innovation in their grammars, the extent to which the rule is obligatory, and so on. Nothing, however, is gradual or quantitative about the occurrence of the innovation. It is merely one more case of rule addition.

This section has dealt with two aspects of phonological change: its *implementation* and its *spread*. Two positions are possible with regard to each of these aspects of phonological change: they are either *abrupt* or *gradual*. There are thus four logical possibilities in viewing the process of phonological change (Wang 1969).

(a) abrupt implementation and abrupt spread
(b) abrupt implementation and gradual spread
(c) gradual implementation and abrupt spread
(d) gradual implementation and gradual spread

Both (c) and (d) are rejected because of the considerable amount of evidence against the gradual, incremental view of the implementation of a phonological change. Possibility (a) is rejected for lack of evidence that the spread of a phonological change is particularly rapid. Some changes spread rapidly, some don't; typically it seems to take at least one generation for a rule to become obligatory. This leaves (b), which accurately summarizes the view of phonological change presented here: the act of phonological change, its implementation, is abrupt, but the spread of a phonological change is gradual.

5.2 THE REGULARITY OF PHONETIC CHANGE

One of the great sustained arguments in historical linguistics concerns the "regularity of phonetic change." The term has been variously interpreted, and much discussion could doubtless have been dispensed with if terms had been defined more clearly. Let us take as a point of departure the precise formulation made by Bloomfield (1933:364):

> [Sound change] affects a phoneme or a type of phonemes either universally or under certain strictly phonetic conditions, and is neither favored nor impeded by the semantic character of the forms which happen to contain the phoneme.

There are at least two versions of what has been called the *regularity hypothesis* (Hockett 1965:186). Common to both is the notion of regularity—that is, phonological change applies to a large number of items in the lexicon and not to just a single morpheme. The first hypothesis, which follows from statements like Bloomfield's (see Postal 1968:235–239), amounts to the very strong claim that phonological change can take place only in purely phonetic environments: no environment of a phonological change can contain a reference to "higher-order" information such as morphological or syntactic class; exceptions to phonological changes, if there are any, occur in strictly phonetic environments. This strongest version of the regularity hypothesis, which we shall call H_1 (H standing for hypothesis), may then be stated as follows:

> H_1: Phonological change is regular, and its environment can be stated in strictly phonetic terms.

Sound changes in conformity with H_1 are, of course, numerous: for example, Indo-European *b d g* > Germanic *p t k* (the environment here is "everywhere"), and Indo-European *p t k* > Germanic *f þ x* except following obstruents. In each case the change is regular and applies throughout the lexicon, and the environment can be formulated without grammatical categories (Noun, Verb, Accusative, Subjunctive) or syntactic structure (Noun Phrase, Verb Phrase). The environments are strictly phonetic, roughly at the level of representation in a generative grammar after the last binary phonological rule has been applied.

A second, weaker version of the regularity hypothesis can be formulated. The basis for this version is the fact that phonological change is not generally found to be limited to single morphemes. If, in other words, a rule is added changing the segment *x* to *y* in a certain diachronic situation, it is not in general the case that *x* > *y* in one word or morpheme, *x* > *w* in another word, *x* > *z* in still another word, and so on with no conditioning factor present. If sound laws, as such changes have been traditionally called, operated like this, there simply could not be a field called comparative linguistics. Rather, we find that in general phonological changes have an across-the-board character, a regularity. They apply across the lexicon, and exceptions to phonological changes fall into three categories: (1) natural subsets of the lexicon (Nouns, Verbs, Adjectives), (2) specific grammatical morphemes and combinations of these morphemes ("first person plural"), and (3) at most a few idiosyncratic lexical items. That is, except for the third category, which

accounts for possible isolated and nonsystematic exceptions to phonological changes, even the exceptions to phonological changes tend to be statable in terms of natural phonological, lexical, or grammatical categories. We are not, however, constrained to hold that the change can be stated in strictly phonetic terms. We therefore relax this requirement and formulate a second hypothesis H_2, which expresses the notion that phonological change is regular in the sense just discussed:

> H_2: Phonological change is regular, but its environment cannot always be stated in strictly phonetic terms.

Dividing the traditional regularity hypothesis into the two versions H_1 and H_2 makes discussion of the entire question easier. It is possible to cite specific linguistic works, or at the least specific instances in different linguists' work, where the term "regularity hypothesis" or its equivalent has been understood either as H_1 or H_2, and it is perfectly possible to accept H_2 while rejecting the stronger claim of H_1.

Let us begin by observing that the regularity of phonological change (whether H_1 or H_2) is an empirical claim: either phonological change is regular in the sense of one of these hypotheses or it is not, and the only way to settle the question is to examine cases of phonological change. If they turn out to confirm either of the hypotheses, then we may regard that hypothesis as correct. The regularity hypothesis in either formulation is an empirical claim which stands or falls in confrontation with the data. A priori arguments serve us badly here.

When we sift through the data for phonological changes, we find, of course, that H_2 is confirmed—confirmed so well, in fact, that it serves as foundation for the branch of historical linguistics known as comparative linguistics. Phonological changes do indeed apply to large classes of lexical items. Often they are context-free, frequently they occur in purely phonetic environments, and they apply across the board without regard for grammatical category in many cases. When, for example, a rule changing *p t k* > *f þ x* everywhere except after obstruents was added to the grammar of the Indo-European dialect that later gave us Germanic, then *every p t k* in the specified environments was affected, and it does not matter whether the word in which *p t k* occurred was a noun, verb, or in the dative case.

The verification of the weak version of the regularity hypothesis is, as we see, a trivial matter. The really interesting question is *why* sound change should be regular in the sense that its domain is greater than a single word. It is not, after all, overwhelmingly apparent that this should be the case (see Dyen 1963). By the view discussed in Section 5.1 that at least some sound changes are gradual and random, it does not follow that sound change should in general affect identical sounds in a large number of words. If phonological change is basically random drift of sounds, why shouldn't *p* in one word drift

off in a direction quite different from that of *p* in another word? Or to push this line of argument to its improbable extreme, why aren't there lots of cases of phonological changes of the following complicated sort?

p > *f* in certain words, *p* > *b* in others, *p* > *w* in still others.
t > *tˢ* in certain words, *t* > *tʰ* in others, *t* > *n* in still others.
k > *g* unconditionally.

Weird cases like this simply do not occur in phonological change (though there are plenty of weird enough phonological changes). If our theory of grammar and language is to account for the facts of historical change, we expect to find in this theory some rationale for the correctness of H_2. The regularity of phonological change in the sense of H_2 does in fact follow from the conception of linguistic change in generative grammar: that speech is the result of an internalized competence, a grammar—a system of rules and a lexicon—and that change consists of alterations in this internalized grammar. In the case of innovation, which is by and large the type of change most commonly referred to as "sound change," a rule is added to this grammar. This rule changes everything that fits its structural analysis. Rules tend to be general—not confined to a single morpheme in the lexicon—so that *every* occurrence of a segment in the designated environment undergoes the structural change of the rule. Phonological changes tend to affect natural classes of sounds (*p t k*, high vowels, voiced stops) because rules that affect natural classes are simpler than rules that apply only to single segments. A rule affecting the natural class /p t k/ is simpler by a feature than the same rule affecting /p t/, and the latter rule in turn is simpler by a feature than the same rule affecting only /p/. That is, a rule applying to:

$$\begin{bmatrix} +\text{obstruent} \\ -\text{continuant} \\ -\text{voice} \end{bmatrix} \qquad /\text{p t k}/$$

is, all else in the rule being the same, simpler than one appyling to:

$$\begin{bmatrix} +\text{obstruent} \\ -\text{continuant} \\ -\text{voice} \\ +\text{anterior} \end{bmatrix} \qquad /\text{p t}/$$

and this rule is simpler than one applying to:

$$\begin{bmatrix} +\text{obstruent} \\ -\text{continuant} \\ -\text{voice} \\ +\text{anterior} \\ -\text{coronal} \end{bmatrix} \qquad /\text{p}/$$

Thus H_2 is adequately confirmed by the data, and generative grammar provides a rationale. It is different with H_1—the strong hypothesis that phonological change occurs only in phonetically defined environments. Nothing in the theory of generative grammar would lend prior logical credence to this claim. In the view advanced here, the class of possible innovations in the grammar of a language is a proper subset of the class of phonological rules. Some phonological rules in natural languages require for their operation grammatical information carried over from the lexicon and the syntactic rules. In English, for example, the rules assigning word stress place stress differently in nouns and verbs, e.g. *cóntent* versus *contént*, *pérmit* versus *permít*. In many languages rules deleting and adding segments apply only to restricted classes such as verbs, nouns, or even subclasses such as strong verbs. Rule 3.13 (discussed in Section 3.3) in the grammars of certain of the Germanic dialects is stated in terms of the grammatical features *Stem-final*, *Past Plural*, and *Past Participle*.

Since this is so, it would be unlikely that every phonological change could be stated in terms of purely phonetic environments. And the empirical evidence bears out this prediction. Cases are not uncommon of changes that occur across the board except in certain morphological environments. In the development of Standard Yiddish from something similar to Middle High German, we find that final unaccented *e*, phonetically [ə], has been lost: *tage* > *teg* 'days', *erde* > *erd* 'earth', *gibe* > *gib* 'I give', *gaʒʒe* > *gas* 'street'. In some cases, however, final [ə] is not lost, principally when the *e* is an adjective inflectional ending: *di groyse shtot* 'the big city', *dos alte land* 'the old country', *a sheyne froy* 'a pretty woman'. A few other final unaccented *e*'s are retained, erratically, but these too are confined to specific morphological environments, e.g. *gésele* 'little street', where *-(e)le* is the diminutive suffix.

The retention of *e* in the adjective endings has nothing to do with a difference in phonetic environment. All schwas were in unstressed position, and there is no phonetic property characteristically associated with adjectives in Middle High German that might somehow account for the loss. We can even find near-minimal pairs containing final unaccented *e*'s that were dropped or retained: *gloyb* 'I believe' : *toybe* 'deaf (inflected adjective)' from Middle High German *gloube* : *toube; meyn* 'I think' : *sheyne* 'pretty (inflected adjective)' from Middle High German *meine* : *schæne*.

Nor is there an explanation in analogy. There is nothing to analogize to in these cases. The simplest conclusion is that the environment of this change is not purely phonetic:

$$5.6 \quad \begin{bmatrix} \text{V} \\ -\text{stress} \end{bmatrix} \rightarrow \begin{Bmatrix} [-\text{next rule}] \,/\, +\ \underline{\quad}\ )_{\text{Adjective}} \\ \emptyset \qquad /\ \underline{\quad}\ \# \end{Bmatrix}$$

(Unstressed vowels are deleted in word-final position unless that word is an inflected adjective. The rule can be stated as applying to *all* unstressed vowels because only *e* [ə] occurs finally under weak stress.)

This, then, is a case pure and simple of phonological change that cannot be stated in terms of purely phonetic features. It is, in other words, a counter-example to the strong form of the regularity hypothesis H_1. A word sometimes used in attempting to account for morphologically conditioned phonological change like this is *functional* (Sapir 1949:262). The notion (in this case) is that *e*'s serving to mark adjective inflections fulfill a necessary function which requires their maintenance, whereas *e*'s in all the other cases can be dispensed with. This is not an explanation for the dilemma but merely a different term to designate it with, for unless "functional" is defined in some precise, noncircular way it cannot be offered as an explanation.

Another instance of phonological change in nonphonetic environments occurs in Mohawk (Postal 1968:245–254), where the sequence [kw] from proto-Mohawk-Oneida sometimes undergoes epenthesis, cf. the pair Mohawk [kewi'stos] : Oneida [kwi'stos] 'I am cold', parallel to a general process of epenthesis in consonant-resonant sequences that breaks up the clusters [wr, nr, sr, tr, kr, tn, sn, kn, tw, sw, kw, sy] by inserting *e*. Certain [kw] sequences, however, do not undergo epenthesis in Mohawk; one is of the same type as the Yiddish example. When the *k* and the *w* in [kw] are, respectively, the first person marker and the first element of the plural morpheme, no svarabhakti (epenthetic) *e* is inserted: e.g. Mohawk [ya'kwaks] : Oneida [ya'kwaks] 'we several exclusive eat it'. There is nothing irregular or sporadic about this: it happens throughout the language in noun and verb prefixes whenever the sequence [kw] means "first person + plural." Like the Yiddish example, it is regular in the sense of H_2 but not H_1. It applies across the board except that it is impeded in a particular morphological environment. (Notice that the existence of morphologically conditioned phonological rules does not force the conclusion that such rules were added in their synchronic form. It is an interesting but yet unproved claim that all such rules are originally innovated as "purely" phonological rules and later restructured to contain morphological information. In the Yiddish and Mohawk cases there is no reason to suppose that the rules discussed were innovated lacking the morphological conditioning.)

On balance it seems unlikely that such morphologically conditioned phonological changes are rare in the world's languages. They do not figure very prominently in formal accounts of historical linguistic development for a variety of reasons. One reason is that they are counter-examples to H_1. A second reason is a certain dullness which attaches to them. Once we have determined that *x* becomes *y* except in the morphological environment *z*, the story is over, and there is little to do but move on to more interesting things. Speculating *why* [kw] did not undergo epenthesis in a particular morphological environment or *why* final [ə] did not drop in Yiddish in adjective

inflectional endings is on a par with speculating why Indo-European $*k^we$ and $*k^wo$ became Indo-Iranian *ča* and *ka*. Usually we simply do not know, though no harm is done by considering possible causes.

One can always devise some ad hoc explanation to save the strong form of the regularity hypothesis when faced with nonphonetic sound changes. Instead of assuming the obvious—that some regular phonological changes take place in environments whose specification requires superficial grammatical structure—one might posit a boundary of some sort (a "plus-juncture") for just these cases. Since many formal boundaries in language do have observable phonetic correlates (word boundary is sometimes realized as pause), one could attribute to the plus-juncture certain purely phonetic characteristics. In this way it is always possible to reduce the original exception to one with a strictly phonetic environment. In the Yiddish example one could assume for Middle High German a plus-juncture (+) that precedes all and only adjective endings and then state the rule of schwa-deletion as: schwa disappears word-finally except after plus-juncture. From *toub+e* 'deaf (inflected adjective)' one would obtain Yiddish *toybe*; from *gloube* 'I believe', Yiddish *gloyb*.

It should be obvious that this is a trick, a gimmick. It is no solution to the problem; it merely provides a simple sign (+) to designate the troublesome cases with. The reason why this is an illegitimate device is that boundaries in natural languages are hardly ever (probably never) consistently realized in some particular phonetic way. In other words, so far as we know, it is a universal that boundaries, whether morpheme, word, or whatever, are optionally realized as null. All experience with currently spoken languages supports this proposition. To postulate for an historical language a kind of unique boundary *always* phonetically manifested in some defined way violates the cardinal constraint in historical linguistics: descriptions of earlier languages must never violate universals that hold for actually observed languages.

The major reason why morphologically conditioned phonological changes have received relatively little attention is that H_1, the strictly phonetic version of the regularity hypothesis, has been held by the majority of the linguists working in the historical field, certainly by those in the Neogrammarian tradition. If one accepts H_1 as a matter of principle, then the question becomes not whether morphologically conditioned phonological changes exist but what other factor or combination of factors accounts for the aberrancy. The following is a typical example (Bloomfield 1933:362–364). Intervocalic *s* from Indo-European is normally lost in Greek: **geúsō* > Greek *geúō* 'I give a taste'. However, in a large number of aorist verb forms we find, apparently, a retained intervocalic *s*: *ephílēsa* 'I loved', *emísthōsa* 'I let', *etímēsa* 'I honored'. This is generally attributed to analogy because aorist *s* is preserved when not intervocalic: *égrapsa* 'I wrote', *épleksa* 'I wove'. In this case the explanation is plausible since there is something of a model for

the analogical reintroduction of *s* in positions where it would have disappeared by regular sound law. Nevertheless, *ephílēsa*, and so on, are counterexamples to H_1, and to save the hypothesis in its strong version we must look elsewhere for an explanation. In the Yiddish and Mohawk examples, analogy is out of the range of reason. Considerations of this kind rule out the strong form of the regularity hypothesis, H_1, but not the weaker form, H_2.

In other cases phonological change can be stated only in terms of a phonological environment that is not purely phonetic. Generative phonology is insistent for many reasons on the difference between abstract levels of phonological representation and phonetic representation. Roughly speaking, the latter is the level of representation after applying the last binary phonological rule (the n-ary rules that fill out the phonetic detail are irrelevant here). Anything higher is more abstract, "deeper" because further removed from the actual phonetic shape. The most abstract level of phonological representation is the string of formatives present as input to the first rule of the phonological component. The striking difference between deep and surface structure has been evident in many of the examples given here, e.g. the phonetic surface form [dəva·i̯n] has a deep structure representation (systematic phonemic, underlying) /divīn/ and intermediate representations such as [divīyn], [divēyn], and [divāyn].

In the light of this hierarchy of phonological representation, the strongest possible form of the regularity hypothesis would be that only surface phonetic structure is permissable to the statement of the environment of a phonological change. This in turn is equivalent to the claim that phonological change consists solely of rule addition at the end of the phonological rules. In this view, every innovation would have to be expressible by adding a rule at the lowest level of phonological representation—the surface level. This is the substance of H_1. In Chapter 3 we examined a number of cases in which this is not true. The only way to express Lachmann's Law in Latin is by assuming that a rule was added *not* at the end of the binary phonological rules but before the rule devoicing obstruents regressively (Rule 3.9). Lachmann's Law thus crucially requires a higher level of representation than the surface phonetic; it requires the representation /agtum/ rather than the surface form [aktum] to give the correct form *āctum* 'having been driven, led'. Without the higher level there would be no way of obtaining the long vowel in *āctum* from surface [aktum] alongside the short vowel in *factum* 'having been made' from surface [faktum].

Notice that it is not claimed here that rules may be added at only two points in the derivation of an utterance—the systematic phonemic and the surface phonetic representations. The claim is not that Lachmann's Law requires the systematic phonemic level of representation for its statement, but only that a rule could not have been added on at the end of the phonological component. We assume rather that the rule was inserted into the grammar of Latin where it applied to derivations somewhere between the systematic phonemic and surface phonetic.

The problem in Mohawk epenthesis discussed earlier and taken from Postal (1968) offers another case of the same kind. It will be recalled that consonant-resonant clusters normally undergo epenthesis in Mohawk, and that some instances of [kw] do so while others do not. In one of the exceptional cases epenthesis was shown to be impeded in a particular morphological environment. Another case of nonepenthesis must be explained in a different way. An example is Mohawk *ra'kwas* : Oneida *la'kwas* 'he picks it up'. Since this is not in the environment "first person + plural," the nonoccurrence of epenthesis must be sought elsewhere. Postal (1968:249) shows that the underlying form of the *kw* in Mohawk *ra'kwas* is /ko/, and that the epenthesis rule applies in the grammar of Mohawk before the rule that converts underlying /ko/ into [kw]. The /ko/ in the underlying form of *ra'kwas* 'he picks it up' is not a consonant-resonant cluster to which the epenthesis rule is applicable; it is a consonant-vowel sequence. Therefore, no epenthesis occurs, and /ko/ later is changed into [kw]. Again, there is nothing irregular or sporadic about this: *every* [kw] from underlying /ko/ comes out [kw], not *[kew]. The explanation is that a rule (epenthesis) is added to the grammar not at the end of the phonological component, where it would operate on surface phonetic forms, but prior to the end so that it operates on more abstract representations (like /ko/ instead of surface level [kw] from /ko/). In this case, the rule converting /ko/ to [kw] belonged to the grammar of proto-Mohawk-Oneida and is several millenia old. The epenthesis rule is only about four hundred or so years old, yet it was inserted into the grammar of proto-Mohawk prior to the /ko/ > [kw] rule. The synchronic ordering of the two rules is (1) epenthesis, (2) /ko/ > [kw]; the chronological ordering is the reverse, (1) /ko/ > [kw], (2) epenthesis.

presumably could be reordering backwards goes from—marked to 4 marks order.

5.3 ANALOGY

Traditionally, historical linguistics has consisted largely of analysis of the interplay between sound change and analogy. Sound change takes place, pattern irregularities may arise; analogy tends to regularize the results.

Sections 5.1 and 5.2 presented arguments against the traditional views that phonological change is reducible to sound change and that phonological change is regular and phonetic in the sense of H_1, the strong form of the regularity hypothesis. The traditional views have several consequences; one is that phonological change not happening to conform to H_1 is forced into categories of change such as analogy and borrowing. The latter categories, in particular analogy, thereby tend to become terminological receptacles devoid of explanatory power—catchalls for irregularities in the operation of "regular sound laws." This has too often been the demeaning fate of analogy in historical work.

Let us formulate the opposing points of view in this way. Traditional historical linguistics has operated within a framework composed of the concepts of sound change, analogy, borrowing, and grammar. Grammar is the account of language structure; it is central. To account for changes in structure (grammar) one appeals to sound change, analogy, and borrowing. As historical linguistics is treated in generative grammar, grammar is enough: "sound change" is grammar change, "analogy" is grammar change, borrowing is grammar change.

The emphasis in this section is on analyzing traditional cases of analogy as part of a process not different in kind from the other types of linguistic change examined so far. In particular, it will be argued that most kinds of "analogy" too are special cases of simplification, in principle very similar to rule reordering, rule loss, and rule simplification proper. The general premise of this section is that analogy in its traditional sense is not some sort of fifth wheel on the wagon, fundamentally at odds with regular diachronic developments like phonological change. In the discussion that now follows, the term "analogy" is to be understood as a cover designation for those instances of change which traditional historical linguistics would have ascribed to analogy.

Analogy is most palpable and most often appealed to in morphology. A typical, uninteresting because transparent, case is the extension of the *s*-plural throughout the nominal inflection of English. The facts are clear. In Old English each noun was characterized in part by its membership in a stem-class: *dæġ* 'day' was an *a*-stem with the nominative plural *dagas*; *caru* 'care' was an *ō*-stem with the nominative plural *cara*; *dēd* 'deed' was an *i*-stem with the nominative plural *dēde*; *tunge* 'tongue' was an *n*-stem with the nominative plural *tungan*. In Modern English, and to a large extent already by Middle English times, the *-(a)s* ending of the masculine *a*-stems has become generalized throughout the nominal system without regard for the original stem class: *cares, deeds, tongues*.

Clearly a simplification has affected at least two components of the grammar: the lexicon and the late transformational rules that attach inflectional endings. In the lexicon of Old English, each noun had, in addition to all the other phonological, grammatical, and semantic information necessary to characterize it, a marker for stem-class. This marker signals the transformational rules for the correct ending. Schematically, then, we would have a set of transformational rules of the following type:

$$dag_{a\text{-stem}} + \text{nominative} + \text{plural} \rightarrow dagas$$
$$caru_{ō\text{-stem}} + \text{nominative} + \text{plural} \rightarrow cara$$
$$dēd_{i\text{-stem}} + \text{nominative} + \text{plural} \rightarrow dēde$$
$$tunge_{n\text{-stem}} + \text{nominative} + \text{plural} \rightarrow tungan$$

Some of the divergence in formation might be accounted for by phonological rules and different base forms, but some morphological marker equivalent to stem-class would still be necessary in at least some cases.

The simplification that has taken place here is a twofold one. In the first place, nouns in Modern English do not require a special marker for stem-class. They are all unmarked in this regard; only exceptional plurals (*sheep, children, men*) need a marker to indicate that they do not undergo the regular rule of plural formation. This is simplification in the lexicon. There is also a concomitant simplification in the number of rules of pluralization: all but the first of the above four rules are deleted from the grammar; the synchronic form of the *a*-stem rule remains and attaches *-z*, giving *cats, dogs, houses*, and so on.

No doubt, many cases of analogy—especially analogical leveling—are of this general type whereby lexical entries become simplified and a rule or set of rules are lost from the grammar while others survive. Other instances of analogy are more whimsical in that no apparent simplification is at work, only a realignment. In attested Old English the plural of *ġiest* 'guest', an *i*-stem noun, is *ġiestas*. The original Germanic nominative plural was *-īz*, which would show up in Old English as *-e*, as in *wine* 'friends'. We should expect the nominative plural of *ġiest* to be *ġieste*, but for all practical purposes *ġiest* has become an *a*-stem noun instead of an *i*-stem noun. This is not lexical simplification in any obvious way, but only the change of a marker; and no compensating simplification occurs in the rules of pluralization: the rule attaching *i*-stem plurals must remain in the grammar to give the correct plural in, for example, *wine* 'friends', *dēde* 'deeds' (though the Early West-Saxon plural *dǣda* shows that *dēd* had become realigned as an *ō*-stem).

Likewise, if someone says today *bring* : *brang* : *brung* instead of correct *bring*: *brought* : *brought*, the analogy and its source are clear, but superficially the speaker has made no formal simplification in his grammar. *Bring* is changed in its lexical entry from [..., − Strong, − Regular] to [..., + Strong], yet the rule for forming irregular weak past tenses and participles remains (*fight* : *fought* : *fought*; *seek* : *sought* : *sought*). One might well argue that "Strong" is in some way simpler than "Weak Irregular," but this claim does require motivation, though it seems intuitively sound. (Frequency of occurrence may have something to do with this type of realignment and with some other kinds of analogy as well.)

Underlying most cases of morphological analogy is a clear argument for simplification. In this view, then, analogy is not different from what is typical of the child's learning of his language. There is disregard of the data in the interest of a simpler account of one's language; there is generalization of a rule beyond its proper domain in the grammar of the older generation. Most of these incorrect creations will be disposed of during maturation, but some may fit the "cut" of the language so well that they become a part of it, especially if the same type of simplification occurs simultaneously in that or closely following generations. The incorrect learning of one generation is the dominant pattern of the next.

The force of this argument leads us to seek parallels between this kind of

simplification and the kinds of simplification found earlier, such as rule loss and reordering. Let us consider the case of Gothic discussed under RULE LOSS in Section 3.3. There it was argued that Gothic once had in its grammar either Verner's Law (Rule 3.12) or its altered synchronic counterpart (Rule 3.13). As long as this rule was present, the surface realizations of verbs in root-final voiceless fricatives included morphs with root-final voiced fricatives: e.g. the principal parts of *kius-* 'to choose' /kiusan kaus kusum kusans/ were phonetically realized as [kiusan kaus kuzum kuzans]. With the loss of the voicing rule, voiced fricatives from underlying voiceless fricatives reverted to their voiceless status: [kiusan kaus kusum kusans]. The crucial bits of evidence were relic forms such as *tigjus* 'decades' (originally from *taihun* 'ten'), *faginon* 'to be happy' (*g* < original *h*, cf. *faheþs* 'joy'), and *alds* 'generation' (*d* < original *þ*, cf. *alþeis* 'old'). These relics, all of which have reflexes of voiced fricatives from original voiceless fricatives via Verner's Law, were disassociated from their sources with voiceless fricatives, hence were restructured in the lexicon with underlying voiced fricatives. There were no synchronic rules linking the pairs cited. The principal parts of verbs, however, were so morphologically cohesive (related by synchronic rules) that they simply reverted to voiceless realizations as in the underlying forms. The few exceptions to the latter statement are all verbs with precarious status and defective distributions, e.g. *aih* 'I possess': *aigum* 'we possess'.

This rule loss introduces a leveling throughout the paradigm of strong verbs. While the rule was still present, some verbs (like *kiusan*) had allomorphs with voiced (*z*) and voiceless (*s*) stem-final fricatives, others had throughout no change in the stem-final fricative: *faran for forum farans* 'to wander', *niman nam nemum numans* 'to take', *mitan mat metum mitans* 'to measure'. The rule loss produces uniformity in accord with the latter type, just as Modern English nouns are uniformly inflected with the *s*-plural except for the handful of relic forms (*men*, and so on).

Rule reordering also brings about a regularization of allomorphic variation, as was pointed out in Section 3.3 in the case of German rule reordering. Before the reordering certain nouns had allomorphs with both long and short vowels : *lop* : *lo:bə* 'praise, praises', *gras* : *gra:zəs* 'grass, of the grass'. Others had only one type of vowel throughout: *bet* : *betən* 'bed, beds', *blu:mə* : *blu:mən* 'flower, flowers', *has* : *hasəs* 'hate, of the hate'. The reordering levels out this kind of variation so that throughout the paradigm nouns have only long or short vowels and not some of one and some of the other.

The tenor of the arguments advanced so far is that traditional analogy, rule loss, rule reordering, not to mention rule simplification proper, are all reflections of a universal process of simplification that ultimately goes back to the child's acquisition of grammar. One might propose a different relation among these processes, namely that *analogy* is the central force and is reflected in such things as rule reordering and rule loss. In this view the change

of *lop* and *gras* to *lo:p* and *gra:s* would be *caused* by analogical pressure from other forms in the paradigm that have long vowels. Similarly, in Gothic the reintroduction of voiceless stem-final fricatives throughout strong verbs would be caused by analogy, the source being the voiceless stem-final fricatives in certain principal parts. Rule reordering in the one case and rule loss in the other would then be mere descriptions of what has happened rather than the prior events.

Consider carefully these two accounts which, let it be noted, are in no way merely different terms for the same thing. There is more at stake here than terminology. We assume, as was discussed at length in Chapter 4, that simplification of grammars is an option always open to the child and that it derives directly from the transmission of language to the oncoming generation. Simplification is the order of the day in the child's acquisition of language, as was proposed and supported by data in Section 4.2. The formal correlates in the grammar of simplification are many: restructuring, rule loss, rule generalization, rule reordering.

The alternative view—that analogy is basic and the other things follow from it—requires us to be very specific about what analogy is and about the rationale for its occurrence. Even more, we must plausibly demonstrate that it accounts for the changes here attributed to rule loss and rule reordering. An enormous amount has been written about analogy, though typically more as an adjunct to an argument than as the object of investigation by itself. Nevertheless, there are studies of analogy per se (e.g. Hermann 1931, Kuryłowicz 1945–1949, Mańczak 1958), and some recent penetrating studies and collections of studies of historical linguistics could serve as textbooks on the process (Benveniste 1948 and 1962, Kuryłowicz 1958 and 1960, Szémerenyi 1960, Watkins 1962).

The traditional theory of analogy is based on the idea of the *proportion* (Paul 1960: Chapter 5). For example, to explain the occurrence of incorrect *brang* for *brought* one "solves" the proportion *sing* : *sang* = *bring* : *x*. In this way the correct but irregular paradigm *bring/brought* gives way to incorrect but regular *bring/brang*.

There are several grave defects in the proportional theory of analogy. First, it is not clear what conditions must be imposed on the forms in the proportion; it is not even clear that conditions *can* be stated which give the right results for each instance of analogical change. For example, *sing* and *bring* agree with each other phonologically in certain ways—among other things they rhyme, so that it makes some kind of sense to put them in the same proportion. Yet how close must the phonological agreement be before two forms qualify as input to a proportion? Must they rhyme? Is it enough to share the last two, or three, or four phonemes in common and in the same order? It seems highly unlikely that satisfactory agreement conditions can be formulated to account for all analogical changes from the world's languages.

Second, no matter what agreement condition is imposed on the items in a proportion, it will not be possible in general to produce proportions for all the forms undergoing changes that one would like to call analogical. Consider the extension of the *s*-plural in English. To show that *caru* 'care' gave up its plural *cara* for *cares* by proportional analogy, one must produce an *a*-stem noun agreeing with *caru* in some close way. And we must do the same for *dǣd* 'deed', *tunge* 'tongue', and the host of nouns that did not originally take a plural in *-s*. This simply cannot be done since, for one thing, *a*-stem nouns ended only in consonants in Old English.

On reflection it should be obvious that defects like these in the proportional theory of analogy cannot be remedied by imposing ad hoc conditions on the proportion. Many just such conditions have been proposed. Vendryes (1925:157) suggested that the formula should be $p : p' = a : x$, where p and p' represent "infinite" quantities. Since no language has an infinite number of lexical items, this condition would make analogical change impossible in every natural language—clearly an unfortunate result for anyone wanting to believe in analogy.

A third, even more serious failure of the proportional theory of analogy is its inability to account for the *regularity* of a very large number of so-called analogical changes. This shows up particularly in cases of rule loss and rule reordering. Given Gothic *kiusan kaus kuzum kuzans* 'to choose' and other strong verbs with a voiceless : voiced alternation, why should *all* of them be analogically realigned with voiceless fricatives? This is not predicted by the proportional theory. Other verbs in the language with no voiceless : voiced alternations present both regular patterns: both voiceless obstruents throughout, e.g. *greipan graip gripum gripans* 'to sieze', and voiced obstruents throughout, e.g. *steigan staig stigum stigans* 'to climb'. Both proportions were, therefore, present in Gothic: *voiceless* : *voiceless* = *kiusan* : *x*, from which we obtain correct *kusum kusans*; but also *voiced* : *voiced* = *x* : *kuzum*, from which we would obtain the nonattested form **kiuzan*. The point is that unless it is integrated into a theory of language in a way not hitherto done, the proportion theory of analogy would lead us to expect *both* kinds of leveling for Gothic: leveling sometimes in favor of the voiceless stem-final fricative and other times in favor of the voiced stem-final fricative. But Gothic does not present us with this at all. There is perfect regularity in that voiceless stem-final fricatives alone survive, and only the few relic forms extant in attested Gothic give the slightest hint of the original allomorphic variation. There is nothing sporadic, idiosyncratic about this; it is as regular as the average phonological change.

Note that the complete regularity of this leveling process is explained correctly by rule loss. The rule converting voiceless fricatives into voiced ones was lost, and all forms previously affected by the rule no longer undergo it. Thus, *all* voiced fricatives produced synchronically by this rule revert to their voiceless counterparts.

Similarly, in the case of German reordering by which *lop* 'praise' became *lo:p* and *gras* 'gras' became *gra:s*, why don't we find irregularities in regard to which vowel, the long one or the short one in the allomorphs of such words, survives in the analogical leveling? And if analogy is proportional, affecting one word at a time in the proportion, why aren't there at least a few nouns still around that have not participated in leveling at all? These are the consequences of the proportion theory of analogy, at least in its current formulations. In point of fact, in German *every* vowel before an underlying voiced obstruent (e.g. [lop] < /lob/, [gras] < /graz/) has become lengthened, so that every noun in the class of *lop*, *gras*, and so on, has undergone leveling in favor of the long vowel. Only relic forms like *weg* [vek] 'away' from earlier *Weg* [wek] 'path', underlying /veg/, in which restructuring to /vek/ has taken place *before* reordering, give us any indication of the earlier situation. (Though some dialects preserve the original ordering.)

Analogy in its traditional interpretation as a proportion gives a plausible account of some of the isolated realignments found in virtually all languages. Analogy as a proportion is a kind of one-shot affair duplicated, if at all, only sporadically elsewhere in the language. But the Gothic and German examples are anything but sporadic and irregular. They demonstrate perfect regularity of rule loss and rule reordering respectively. Only when we think of each separate verb and noun form in these languages as solutions to a proportion does the regularity become anything to wonder about. Similarly, as simplification both in the lexicon and in the transformational component, the generalization of the *s*-plural in English is not remarkable. But if it proceeded proportion by proportion, word by word, we would reasonably expect to have more archaic forms like *men* and *brethren* than we actually have.

Consider again the argument tentatively put forth earlier—that analogy should be taken as basic with simplification its consequence. No matter what interpretation we attach to analogy, this relation is bound to be unsatisfactory. In the first place, the proportional theory of analogy seriously fails to account for changes easily explained as simplification. This we have seen. Second, if we accept a much stronger conception of analogy as an irresistible force that *requires* change each time its conditions are met, it is clear that counter-examples can easily be found. If, in other words, we should entertain the claim that each input to an analogical proportion *must* undergo change, we would easily be able to falsify our claim. The noun *foot* still has the plural *feet*, the noun *child* still has the plural *children*; these are counter-examples to analogy as an irresistible force. If we accept the weaker interpretation of analogy as something that merely points the direction of possible change, then analogy becomes superfluous because simplification is enough. This weaker version of analogy would claim that *foot* and *child*, if they give up their old plurals, will become *foots* and *childs*. But since this is already predicted by simplification, analogy is unnecessary. For this reason, as well as

for the others advanced in this section, we reject analogy. Grammar and simplification are enough.

In emphasizing the regularity of certain changes customarily attributed to analogy but here attributed to simplification, we should not overlook the fact that simplification in the sense used here sets the stage for queer realignments that dot every language. Latin rhotacism provides us with a simple example. A rule changing $s > r$ intervocalically was added to the grammar of Latin sometime prior to the 4th century B.C., giving rise to alternations such as *amīcus magnus* : *amīcōrum magnōrum* 'a great friend, of great friends'; *genus* : *generis* 'kind, of the kind'; *honōs* : *honōris* 'honor, of honor'. The latter alternation is found in later Latin as *honor* : *honōris*. The forms for 'a great friend, of great friends', and so on, remain, however, with their *s* : *r* alternation. What has happened is that the underlying form of 'honor' has changed from /honos/ to /honor/. This case of restructuring is trivial in that the underlying representation of this one word (and a few others such as *ebur* 'ivory') has changed. The rhotacism rule remains in the language, producing *s* : *r* alternations as before, but restructuring has removed this one word from the domain of the rhotacism rule. Such minor lexical changes are absolutely compatible with the concept of change adhered to in generative grammar, and one need not appeal to analogy.

In the same vein consider an example from German. Middle High German had several patterns of noun pluralization, among them (1) addition of *-e*, e.g. *tac* 'day' with the plural *tage*, and (2) umlaut of the root vowel with addition of *-e*, e.g. *kraft* 'power' with the plural *krefte*. Both rules of pluralization are retained in modern German, cf. *Tag/Tage* and *Kraft/Kräfte*. The noun 'tree' obeyed rule (1) in Middle High German, rule (2) in modern German: *boum/boume* versus *Baum/Bäume*. There is no obvious case for simplification that can be made here; both rules of pluralization are carried along unchanged, the lexical entry for 'tree' is no simpler now in German than it was in Middle High German. There has been an idiosyncratic realignment of this particular noun.

5.4 EXCEPTIONS TO PHONOLOGICAL CHANGES

This general line of consideration can be pursued a bit further. Traditionally, one disadvantage of a strong form of the regularity hypothesis (such as H_1) is the existence of all kinds of exceptions to otherwise regular phonological changes. Included here are not only morphologically conditioned changes like the loss of final schwas in Yiddish, which have excited little interest, but more particularly the bewildering array of exceptions and irregularities that have come to light. Romance linguistics is an especially rich mine of such complex linguistic situations, and it is no wonder that the most convinced and vocal skeptics of the regularity hypothesis traditionally have come from Romance philology (Schuchardt 1885, Hermann 1931).

Examples are rife: paradigmatic resistance to phonological change, the interference of folk etymology in regular change, the avoidance of homonymy. An instance of the latter type occurs in certain French dialects where intervocalic *r* has become *z* regularly except for certain words such as *frères* 'brothers' and *oreille* 'ear'. This exception has been attributed (Lerch 1925: 80–82) to a striving to avoid homonymy, to avoid falling together with *fraises* 'strawberries' and *oseille* 'sorrel'. This kind of active participation by the speaker in the processes affecting his language is taken for granted in the work of many Romance linguists, precisely because, one suspects, Romance dialectology has turned up so many exceptions to supposedly regular sound laws. The notions of "therapeutic change" and "lexical pathology" come to mind here, concepts which are exemplified most vividly in the work of the French-Swiss scholar Jules Gilliéron. (Cf. Gilliéron 1915, 1918, also Malkiel 1967.)

Anyone familiar with dialect studies over an extensive language area is not surprised by exceptions to phonological changes that are for the most part regular. Do such irregularities falsify the theory of phonological change proposed here? The answer is No. In this concluding section we shall look into some of the ways that these irregularities may be accounted for in a theory of linguistic change compatible with generative grammar.

Every theory of grammar must be equipped with some way of marking exceptions to general rules. In some languages there is a division between native and nonnative morphemes; typically the latter do not undergo a rule or set of rules affecting the native portion of the lexicon. In Finnish (Harms 1968:120) proper nouns with a single noninitial stop are not subject to a certain rule, which we shall designate as Rule *x*. We account for this in the grammar in the following way. A redundancy rule uses the feature [+ Proper] to state what is special or aberrant about proper noun lexical morphemes. In this instance we would have the redundancy rule:

$$5.7 \quad \begin{bmatrix} +\text{Noun} \\ +\text{Proper} \\ +\text{obstruent} \\ -\text{continuant} \end{bmatrix} \rightarrow [-\text{Rule } x] \,/\, [-\text{obstruent}] \_\_\_\_ \text{V}$$

This states that any noninitial stop in a lexical item fitting the structural analysis of Rule 5.7 is marked additionally as "minus Rule *x*," which by convention prevents such items from undergoing Rule *x*. Similarly, morphemes foreign in a language and therefore exceptions to certain rules will be marked [+ Foreign] in the lexicon, and we will have a redundancy rule of the form:

$$[+\text{Foreign}] \rightarrow \begin{bmatrix} -\text{Rule } x \\ -\text{Rule } y \\ \cdots \end{bmatrix}$$

indicating that foreign morphemes do not undergo Rule *x*, Rule *y*, and any other rules (...) not applicable. (On the treatment of exceptions see Chomsky and Halle 1968:172–176.)

In the cases discussed above exceptional morphemes constituted some small portion of the lexicon. Often, only a single item in the vocabulary is aberrant, usually inexplicably. It just happens not to undergo some rule or rules. English has a laxing rule (Chomsky and Halle 1968:180–181) that laxes the underlying tense vowel in the second syllable of *divinity*, *serenity*, *profanity*, and a large number of other words. An exception is *obese/obesity* in most dialects of English: the laxing rule would normally lax the second vowel in *obesity*, yielding *[obɛsətiy], whereas [obiysətiy] is the usual form. This is simply an idiosyncratic property of this morpheme, and to account for it we mark *obese* as [– Rule *x*], where *x* is the Laxing Rule in the lexicon. By convention every segment of a formative does not undergo a rule for which the lexical morpheme is marked "minus that rule," hence the tense /ē/ underlying the second vowel of *obesity* will not be laxed to [ɛ].

Similarly, we can use such a marking convention in historical linguistics to account for innovations (rule additions) that sporadically and idiosyncratically pass over individual items. In the case of *fréres* and *oreille* mentioned earlier, which would normally have become **frezes* and **ozeille*, we simply assume that these two items were (for reasons not relevant in the immediate discussion) marked as [– Rule *x*], where *x* is the rule converting *r* to *z* intervocalically. Hence, these two items do not undergo the rule converting intervocalic *r* into *z*.

We have not endeavored here to answer the separate question of *why* these two words remained exceptions to a regular phonological change. As a part of sociolinguistics the question would be well worth pursuing: there is no telling what one might find out about ears and brothers, not to mention strawberries and sorrels, in French peasant society. Our concern here, however was rather more astringent: to show that generative grammar has a well-motivated way for dealing with lexical exceptions to rules. Since every language has exceptions to rules, an adequate theory of grammar must be able to cope with them, and one way is the "minus rule feature" as illustrated above. But it is not the role of grammatical theory to explain exceptions to general rules—to explain, in other words, why general rules are not even more general.

A more interesting case of the same kind also comes from Romance linguistics: we might refer to it as a phonological change brought about by morphology (Malkiel 1968). In Old Spanish the behavior of the Latin medial clusters -RG-, -LG-, -NG- before front vowels is aberrant: *ǵ*, a palatalized *g* (-RǴ-, -LǴ-, -NǴ-), shifts to *z*, which was either an affricate [dz] or a fricative [z]. Examples: ARGILLA 'potter's clay' > *arzilla*; GINGĪVA 'gum' > *enzía*; ĒR(I)GERE 'to raise' > *erzer*. This is odd in the general framework of Romance linguistics since the second element of similar

medial consonant groups generally remains unchanged: MORDĒRE 'to bite' > *morder*; VENDERE 'to sell' > *vender*. Word-initial Ǵ- never gives *z*- in Old Spanish; cf. GELĀRE 'to freeze' > *elar*, GERMĀNU '(half-) brother' > *ermano*, where G- before front vowels has disappeared.

None of the various attempts to account for this has been widely accepted or even made very plausible. Malkiel (1968) has presented evidence from various sources for the following explanation. A model for the alternation of *g* (before nonfront vowels) and *z* (before front vowels) is provided by the common verb DĪCŌ, -ĔRE 'to say', which has the Spanish reflexes (indicative) *digo*, *dize(s)*, ..., (subjunctive) *diga(s)*, and so on. The pattern of this phonological alternation has spread beyond the original verb. Thus, in the verb FACIŌ, -ERE 'to do', Old Spanish replaces regular **faço*, *faz(es)*, ..., **faça* by *fago*, *faz(es)*, ..., *faga*, which has been remodeled along the line of *digo*, *diz(es)*, ..., *diga*. The realignment has spread even further, producing unetymological *g* in other words, for example *oiga* in place of *oya* < AUDIA(M) 'to hear (l. pres. act. subj.)'.

The original rule, accounting for the *g* : *z* alternation in *digo*, had, we assume, the status of a "minor rule" in the language. The environment of the rule was simplified, extending its domain to *fago* and other verbs in the language, especially *-ngo* verbs (FINGŌ, PANGŌ, TANGŌ); and the end result of a progressive series of simplifications in the environment of the rule was a rule *not* restricted morphologically but quite generally applicable to any -NG- before front vowels. This rule, now independent of morphological class, is generalized to affect -RG- and -LG-.

Nothing in this richly varied problem conflicts with our picture of phonological change in generative grammar. It is fundamentally a case of innovation followed by simplification, doubtless made more complex by rule borrowing within Romance and muddied still more by other trends in the language(s). This kind of morphological conditioning of and interference with phonological change is embarrassing only if we insist on the strong version of the regularity hypothesis (H_1). If, as has been argued here, H_1 is not observed, then there is nothing particularly upsetting about this type of change.

It has been proposed that the paradigm of change suggested by Malkiel is even more general than has been realized in the literature of historical change. Let us consider, for a moment, the difference between *major rules* and *minor rules* (cf. Lakoff 1965). Major rules in phonology are exemplified by most of the rules stated throughout this book: they apply automatically to every form unless that form has been marked "minus the rule" by a morpheme feature such as [+ Foreign], [– Rule *x*], and so on, as discussed earlier. A minor rule obeys the convention that no form undergoes it unless specifically marked. (Cf. Lightner 1968 on minor rules in Russian phonology.)

In the case from Old Spanish a minor rule applied only to DĪCŌ, thus making an exception of this tiny part of the lexicon. Subsequently a greater

number of morphemes came to be marked as having to undergo this minor rule, e.g. FACIŌ and AUDIA(M), and finally the rule became a major rule applying across a large part of the lexicon.

Wang (1968, 1969) has proposed that in general, phonological change begins in the way indicated by the Old Spanish example. In other words, change starts as a minor rule, making exceptional those items in the lexicon that it affects. The rule extends its domain of application to ever larger portions of the lexicon, eventually becoming a major rule in the grammars of subsequent generations. Exceptions to "sound laws" are then the handful of lexical items not reached by the rule, i.e. those items marked "minus the rule" for the rule in question. (See Wang 1969 for a number of additional conjectures on the cause of irregularities in sound changes.)

However, against the view that phonological changes are initiated as minor rules there is evidence of the following sort. It is possible to show that most minor rules in languages are the synchronic relics of once general phonological rules. There is in modern English a minor rule that accounts for voiceless/voiced alternations as in *leaf/leaves*, *bath/baths*, *house/houses*, and a relatively few other items. To account for this assume that such forms end in underlying voiced fricatives (for example, *house* /hūz/) and that they are marked as having to undergo a minor rule devoicing final fricatives, thus /hūz/ > [hūs] and eventually [haws]. This is clearly a minor rule in English phonology, yet in Old English the rule was completely general: all lexical items were subject to it without exception. Similarly, there was in Gothic a minor rule affecting the *t* of the past tense morpheme (compare *hafts*, *salboþs*, *salboda*, where *t*, *þ*, and *d* are allomorphs of past tense /t/). This was a minor rule of Gothic phonology affecting only *t*'s belonging to certain grammatical morphemes, yet it is the impoverished synchronic residue of part of Grimm's Law—the addition of a completely general rule. This type of data argues against the claim that regular phonological changes originate in the addition of a minor rule. The other way around seems more likely.

In conclusion, let us briefly consider a point often made in the context of these and similar discussions: that the heuristic advantages derived from observation of strict phonetic regularity of phonological change make the strong form of the regularity hypothesis worth retaining even at the expense of its "slight" incorrectness. In other words, we should act *as if* every phonological change were at most phonetically conditioned. The underlying sentiment behind this procedure is that to cease observing H_1 would throw us back to the grim days of Bopp and Rask when no holds were barred in describing phonological change, when one was not constrained to feel that if *p* > *f* in one word, it should become *f* in all phonetically similar words.

However useful this notion may have been at early stages in the development of historical linguistics, it has nothing to do with the truth value of H_1 and should be disposed of once and for all. In discussing as separate entities

H_1, the strictly phonetic version of the regularity hypothesis, and H_2, the weaker claim about phonological change regularity, the main purpose was to demonstrate that a linguist can accept one hypothesis (H_2) while rejecting the other (H_1). This in fact was done here. We have accepted H_2 and even pointed out specific reasons in generative grammar why H_2 should be true. But various kinds of data were produced to falsify H_1, and there is no reason to accept the constraint on change that H_1 embodies.

This does not, however, open the field to wild orgies of unbridled speculation. The more general a rule is, the more highly valued the grammar containing that rule is in the evaluation of grammars. A rule that specifies a change in a purely phonetic environment is higher valued than a rule carrying out the same change in the same environment but now modified by a specification [− Class *x*], where "Class *x*" is a nonphonetic specification such as [+ Noun], [+ Adjective], or [+ Plural]. If all else is equal, the first formulation of the rule is to be preferred over the second.

In short, we try to render the simplest account of the facts. If a change has a purely phonetic environment, the simplest account involves writing a rule with a purely phonetic environment. If the change cannot be stated in purely phonetic terms, we still render the simplest account we can. This may require us to write a rule whose structural analysis contains some morphological features (as in the loss of Yiddish final schwa); or to order the rule in the grammar so that it operates on an abstract phonological representation (Lachmann's Law); or to write a minor rule applying to only a small part of the lexicon; or possibly to write a major rule to which several lexical items are marked as exceptional.

This is all merely a complicated way of saying that historical linguists do what they are supposed to do: describe change. A wide array of evidence now shows that phonological change takes place in environments both phonetic and nonphonetic. To describe change we cannot observe a dictum requiring us to make the environment of every phonological change strictly phonetic. This is just the way things are.

SUPPLEMENTARY READING

Bloomfield, Leonard. 1933. *Language*. New York: Holt, Rinehart & Winston, Inc. Chaps. 20–23.

Hermann, Eduard. 1931. *Lautgesetz und Analogie*. Berlin: Weidmannsche Buchhandlung.

Hockett, Charles F. 1958. *A Course in Modern Linguistics*. New York: The Macmillan Company. Chaps. 42–54.

Kiparsky, Paul. 1965. *Phonological Change*. Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Postal, Paul M. 1968. *Aspects of Phonological Theory*. New York: Harper & Row, Publishers. Chaps. 10–15.

Wechssler, Eduard. 1900. *Giebt es Lautgesetze*? Halle: Max Niemeyer.