1

PROLOGUE

To define historical linguistics is not taxing. One might say that historical linguistics is the study of all aspects of language development through time, or that historical linguistics is the investigation of language change. The average linguist would accept such definitions without much disagreement, perhaps adjusting them here and there in the direction of his private interest in the field. But to define the relation of historical linguistics to general linguistics is not an easy or obvious matter. This is a seminal question having no answer to which all linguists readily assent. Nor is it obvious why historical linguistics is of interest to the linguist, who can do linguistics without caring in the least what historical linguistics is all about.

This has not always been so. There was a time when the position of historical linguistics was quite well defined and when its study was as much a matter of course as that of any other reasonable subject. The reason was simple:

1

main stream linguistics *was* historical linguistics—what you did if your interest was language rather than politics or the cavalry. The place was Europe, particularly Leipzig in Germany, and the time was the last quarter of the nineteenth century, a period of enthusiastic and valuable activity (valuable even today) in the still nascent discipline of historical linguistics. In that cultural and historical context to have asked one of the great linguists of the time, say Karl Brugmann or Hermann Paul, why he was involved in historical rather than general linguistics would have elicited for the sophisticated questioner a puzzled look but not much of an answer. Historical linguistics was where the action was: if you studied languages, you were involved in historical linguistics, and few scholars engaged in linguistic research stopped long enough to wonder whether their work properly belonged to historical linguistics or to the larger field of general linguistics, or indeed to ask what the difference was.

Only an intellectual climate like this could have produced the famous statement made by Hermann Paul in defense of the title of his book *Sprachgeschichte* (*Principles of the History of Language*), published in 1880: "The objection has been raised that the historical approach to language is not the only scientific method available for the study of language. I am forced to deny this" (Paul 1960:20). Such simplicity is not possible now. Few linguists today would care to argue for the advantages of making historical linguistics coterminous with linguistics in general.

Though most of our predecessors in the late nineteenth century (including both Neogrammarians and linguists like Hermann Collitz and Hugo Schuchardt who were not Neogrammarians) merged what we today might call *descriptive* and *historical* practice, at least one of them, Ferdinand de Saussure, came to perceive clearly the fact that the two disciplines are different in several crucial respects. It was Saussure who first insisted on the distinction between what he called *synchrony* (the study of language in its static states) and *diachrony* (the study of language in its evolutionary stages). And the degree to which we now observe the separation of the two betokens the extent to which modern linguistics continues to profit from certain of Saussure's insights. He wrote:

For a science concerned with values the distinction [between synchrony and diachrony] is a practical necessity and sometimes an absolute one. In these fields scholars cannot organize their research rigorously without considering both coordinates and making a distinction between the system of values per se and the same values as they relate to time (Saussure 1959:80).

As with Saussure's writings in general one may quarrel with certain aspects of this statement, or subject it to reinterpretation within whatever theory of language one subscribes to. By and large, however, most linguists today

recognize the necessity of the distinction between synchrony and diachrony, though as Roman Jakobson perceptively remarks, "The uniting of static and dynamic processes in language is one of the fundamental dialectic antinomies which serve to characterize the concept of language. The dialectic of language development cannot be understood without observance of this antinomy" (1931:267).

Saussure's sharp separation of the two aspects of language study continues to have many implications for linguistics. One of its effects was to spoil the easy answer which Hermann Paul and others of his time might have given to the question "Why be concerned with historical linguistics?" After Saussure it was necessary to make a choice, so to speak: if you were interested in historical languages you studied linguistics, and you could if you liked study historical linguistics.

The unfortunate result was a deflection of historical linguistics from the center of the stage in linguistic research—a state of affairs which has continued to the present. Most of the considerable work in structural linguistics done during and since the second quarter of the twentieth century has been primarily devoted to problems of synchronic analysis: refinement of the notion "phoneme," problems of phonological and morphological analysis, immediate constituent cuts, the adequacy of phrase-structure grammars, the role of transformations, and so on. Many European scholars continued to devote most of their time to historical linguistics, but the emphasis in mainstream American linguistics lay very much more on the side of synchrony. The main reason for this, of course, was that under the influence of Boas, Sapir, and Bloomfield, American linguists had been primarily concerned with describing living languages, especially the native American languages, and they required the development of a body of analytical techniques for the synchronic description of language. If diachrony came into the picture, as it did in Bloomfield's Algonquian studies and in much of Sapir's work, so much the better; but synchrony had the prior claim.

One should not leap to the conclusion that historical linguistics in America had necessarily to suffer on account of this shift of attention. On the contrary, the predominant feeling in the forties and fifties was that historical linguistics had everything to gain from assimilating the sophisticated tools being devised for synchronic analysis. It was, as Hall says regarding one branch of historical linguistics, "the goal of some present-day workers to return to comparative reconstruction and to revivify it with an infusion of descriptive (synchronic) analysis" (1950:6). A number of well-known papers illustrate this: Twaddell (1938), where the notions of allophone and phoneme are related to scribal practice in the attempt to explain a long-standing problem in the orthographic representation of High German umlaut; Hall (1950), which is a case study in careful application of the comparative method with constant reference to what earlier would have been called "the phonemic principle"; Hoenigswald (1950), where the problem of comparative reconstruction is shown to

parallel the problem of assigning phones to phonemes on the basis of phonetic similarity and complementary distribution. Hoenigswald (1960) represents in some ways the culmination of such trends. *Language Change and Linguistic Reconstruction* summarizes the directions historical linguistics had been moving under the aegis of American structural linguistics, and it codifies a great deal of what seemed to be best in the emergent techniques.

Meanwhile, in Europe, work in historical linguistics had gone right on, though on lines hardly parallel to those developing in America. Some linguistic scholars had remained outside the structuralist schools altogether. Linguists of the Prague Circle were not primarily interested in historical linguistics, though some of them, notably Roman Jakobson, made important contributions. Jerzy Kuryłowicz continued to publish significant work dealing with theoretical problems in historical linguistics, as did Émile Benveniste; and all over Europe scholars carried on research in historical linguistics either guided by Neogrammarian or "idealistic" precepts or influenced by Prague (e.g. André Martinet and Jean Fourquet).

But all of this has not brought us any closer to answering the second of our initial questions: what is the position of historical linguistics? Or more directly: why should anyone be concerned with historical linguistics? Merely observing that linguistics since the Neogrammarians has not been identical with historical linguistics does not provide a *raison d'être* for the latter. The observation that historical linguistics has been revitalized by new insights from synchronic linguistics is not necessarily calculated to win students for historical linguistics.

More than this, our question seems to gain in seriousness the further away we move from the certainties of the nineteenth century, where so much, not only questions in linguistics, seems to us to have had such easy answers. Throughout the twentieth century linguistics has expanded its horizons (occasionally to its own disadvantage) to include or touch on an ever increasing number of previously excluded subjects: anthropology, psychology, mathematics, sociology, and so on, and we have come to expect, perhaps too optimistically, that linguistics will continue to grow in relevance to mankind's problems and the human condition.

So too with historical linguistics. It is not quite sufficient today to assert that one should study historical linguistics because it is intrinsically interesting (though no doubt that is sufficient reason for the majority of scholars busily engaged in historical linguistics). Similarly, it is not enough to claim that historical linguistics is useful because it gives us new perspectives about present-day languages and how they came to be that way.

The purpose of this book is to present historical linguistics as it is understood and practiced by linguists committed to the conception of language and grammar implicit in the theory of generative grammar. The book's larger purpose, however, is to point out some of the ways in which the theory of generative grammar has made historical linguistics more relevant to general

linguistics. We will be able to discuss such matters critically only after we have examined the form of historical linguistics as it appears within the theory of generative grammar. Suffice it to remark here that historical linguistics and the evidence of historical change have a good deal to say about the ultimate shape of our model of what a speaker knows in order to speak his language.

In other words, we shall see how historical linguistics stands to profit from the new insights into language that have come with the development of generative grammar. But the gains are not all one-sided: the theory of generative grammar too is enriched by the inclusion of data from historical linguistics. We are, in a sense, riding out a pendulum swing from earlier positions in which historical linguistics was receiving much and returning little vis-à-vis linguistic theory. Today we find that material from historical linguistics figures prominently in many discussions principally concerned with questions of synchrony, and often the historical data are crucial in deciding questions about the form of grammar rules (cf. Bach 1968 and Kiparsky 1968b). In the 1950's such cooperation between synchrony and diachrony would have been almost unthinkable.

Our more immediate concern, however, is not so much to justify the existence of historical linguistics or to emphasize its pertinence in regard to general linguistics as it is to examine historical linguistics from within the theory of generative grammar. What do we mean when we say that language changes? How does it change? What changes? In particular, what is sound change, and what is the mechanism by which it takes place? Is sound change regular? What is analogy within generative grammar? What can we say about syntactic change? Does generative grammar have any implications for internal and comparative reconstruction? Why does change occur? What, if anything systematic, do scribes write?

But in delving into these and related subjects, let us not lose sight of our main goal in historical linguistics—to say all that we can about the processes that take place in language through time and space. It is emphatically not our job to provide a number of gimmicks which somehow make language change look easy, nor is it even, primarily, to show that a given theory of language and grammar renders a better accounting of the facts of language change than do other theories. Of course, it is impossible to discuss any set of linguistic data in a vacuum; data do not explain themselves, only an associated theory can explain them; and success in explaining the data is directly proportional to the relative correctness of our theory of language. We cannot discuss language change without discussing the ways in which our theory of grammar gives us richer insights into the nature of change than do other theories. This is proper, but let us not forget that our primary goal is to deepen understanding of the mysterious phenomenon of linguistic change.

First, we must state clearly the theoretical position from which we shall be examining the data from the history of languages. What are the goals of

linguistic theory, and how are we to account for observable facts about language?

SUPPLEMENTARY READING

Hoenigswald, Henry M. 1960. *Language Change and Linguistic Reconstruction.* Chicago: University of Chicago Press.

Lehmann, Winfred P. 1962. *Historical Linguistics: An Introduction.* New York: Holt, Rinehart & Winston, Inc.

Lehmann, Winfred P., ed. 1967. *A Reader in Nineteenth-Century Historical Indo-European Linguistics.* Bloomington: Indiana University Press.

Meillet, Antoine. 1921. *Linguistique historique et linguistique générale.* Paris: Champion (= Collection linguistique publiée par la Société de Linguistique de Paris, 8).

Paul, Hermann. 1960. *Prinzipien der Sprachgeschichte* (6th ed.). Tübingen: Max Niemeyer. Available in English as *Principles of the History of Language,* trans. Herbert A. Strong. London: Longmans, Green & Co. Ltd., 1891.

Pedersen, Holger. 1962. *The Discovery of Language,* trans. John W. Spargo. Bloomington: Indiana University Press.

Waterman, John T. 1963. *Perspectives in Linguistics.* Chicago: University of Chicago Press.

Whitney, William D. 1883. *The Life and Growth of Language.* New York: D. Appleton and Company.

2

BACKGROUND

Most branches of science that obtain their data from the observation of human behavior draw a distinction between what the human organism *knows* and what it *does*. In linguistics this is referred to as the difference between the *competence* and the *performance* of the speaker-hearer. Competence is the intrinsic, largely unconscious knowledge underlying our ability to speak and to understand what is spoken. Performance is the way this intrinsic knowledge is applied in a given case. The study of competence may be regarded as the study of the potential performance of an idealized, "perfect" speaker-hearer. The actual performance of a speaker-hearer is affected by nonlinguistic factors such as distraction, memory limitation, and emotions. The difference is crucial to an understanding of what a generative grammar is supposed to represent, and such understanding in turn is necessary for the material on historical linguistics to be presented in the remainder of the book.

7

2.1 COMPETENCE AND PERFORMANCE

Let us consider a nonlinguistic example. Suppose we wished to formulate a theory of arithmetic ability based on observation of educated adults. For simplicity's sake we shall deal with addition only.

The starting point toward such a theory of addition is the observation of humans in the act of doing sums. We note that our subjects can add certain integers with no apparent difficulty, e.g. $9 + 8 = 17$, $3 + 2 = 5$, that they have some slight trouble with larger pairs of integers, e.g. $128 + 52 = 180$, and that for still larger integers they resort to pencil and paper to keep track of intermediate results, e.g.

$$\begin{array}{r} 658 \\ +\,1798 \\ \hline 2456 \end{array}$$

Observations such as these are the raw material—the *primary data*—available to us in developing a theory that correctly accounts for the exact form and substance of the underlying knowledge involved in addition. By "correctly" we mean that our theory not only gives the right sum for any pair of integers but also is the psychologically correct version of this intrinsic knowledge.

In this sense, certain theories of arithmetic competence in humans are clearly wrong—for example, the theory that the human brain has large addition tables stored within its cells, and that the competence underlying addition consists of a table-lookup procedure. Given any two integers, we just look through these tables until we find the one appropriate for the two integers and read off the result. This theory might be taken seriously at least in the initial stages of our work since it could be made to predict the right answers. Moreover, it would reasonably account for one aspect of the primary data: that our subjects tend to take more time carrying out addition as the numbers become larger. Under our tentatively proposed theory the longer time could be attributed to the need to look through more of the internalized tables.

But this theory of competence is falsified (i.e. shown to be incontrovertibly wrong) by at least one crucial piece of evidence: the competence of human beings to produce an infinite number of arithmetic results. There is simply no upper limit to the sums we can in principle produce. No matter what two integers we are supplied with, we can calculate their sum in a finite amount of time. Suppose someone were to claim that the upper limit to the sums a human being can produce is 145,987,823,975,576. I could merely add 1 to this, obtaining 145,987,823,975,577. Since the brain, though it has immensely many cells, does not have an infinite number, and since human beings can in principle generate a number of sums in excess of any finite number, it is

obvious that we can produce more sums than can be stored in the brain. Our rather simple-minded table-lookup theory of arithmetic ability must be wrong.

Various theories that do not entail infinite brain capacity can be envisioned. For example, we might assume that addition is based on a combination of two basically simple operations: one is a table-lookup within a finite, even small, table; the other is a recursive (i.e. repeatable) process requiring a small amount of temporary storage capacity. We assume that the brain has in permanent storage an addition table of the form $m + n$, where m and n are integers from 0 to 9. Addition of two integers each of which is less than 10, say 6 and 3, is carried out by table-lookup within this small table (small since it has only 100 entries). Addition of integers larger than 9 consists of a lookup in this table, temporary storage of this result as well as of the "carry," repeated table-lookup, and so forth with repeated application of these two processes. This procedure, which resembles what children are taught to do in school, can best be illustrated with a simple example.

Suppose we add 29 and 34. First we obtain $9 + 4 = 13$ by table-lookup since both 9 and 4 are integers less than 10. We store the 3 in our result cell in the rightmost position and store the 1 in a temporary "carry" cell. This corresponds to the initial step "9 plus 4 equals 3 carry 1" in the language of the elementary schoolroom, or

$$\begin{array}{r} 1 \\ 34 \\ +\,29 \\ \hline 3 \end{array}$$

Next we obtain $3 + 2 = 5$ by table-lookup, and this 5 is added (via table-lookup) to the 1 left over from the previous addition, yielding 6. This result is then entered in the position second from left in the result cell, which now contains the final answer 63.

Such an account of arithmetic facility is not incorrect in the way that the theory of mere table-lookup was. It assumes only a finite amount of storage capacity, yet can reasonably account for our ability to produce arbitrarily large sums by finitely many applications of simple processes. We must ask ourselves, however, whether this is an account of competence in the technical sense of that word. Our proposed theory is free, to be sure, of gross performance factors such as whether the subject shouted or whispered his answers when asked to add, whether he mumbled them, whether he made a mistake when momentarily distracted by a pretty girl. All these things are clearly performance factors: they affect given acts of addition, but they have nothing to do with the intrinsic knowledge of a person who has learned to add.

But in a more interesting sense it is by no means obvious that our theory is one of competence and not of performance. Is it not possible and even likely

that we have a quasi-performance model, stripped of the grossest performance factors but in reality far removed from what actually goes on psychologically when someone adds two numbers? Haven't we done what is usually done as the first step in the history of any science—namely, haven't we simply equated appearance with reality, assuming that what we see on the surface is really what *is*? All we can assert is that our account meets the level of *observational adequacy*: it describes the data on which it is based (Chomsky and Halle 1965:99). Let us now briefly consider another possible candidate for a theory.

We might assume that the correct account of the tacit knowledge underlying addition is more abstract, less tangible, than assumed in the previous theory. We might assume further that arithmetic involves repetition of a procedure which we may call the "successor operation." The "successor" of an integer is the next larger integer: the successor of 2 is 3, the successor of 298 is 299. Therefore, given two integers, for example 28 and 52, we may propose that their sum is obtained by means of 52 consecutive applications of the successor operation to 28: $28 + 1 = 29$ (one application), $29 + 1 = 30$ (two applications), . . . , $78 + 1 = 79$ (fifty-one applications), $79 + 1 = 80$ (fifty-two applications and halt). This account also requires only a finite amount of storage capacity and permits the addition of any two integers in finitely many steps.

Again, we note that this theory meets the level of observational adequacy and that it is not a performance model in the trivial sense of performance: it makes no predictions about how a particular act of addition is carried out (i.e. whether accompanied by grunting, whether pencil and paper are used, whether the subject said "81" when he meant to say "80"). This account of arithmetic ability is more abstract, further from the primary data of observation, from the surface aspect of addition, and closer to an account of what really takes place when we add two integers.

Both theories lay claim to our consideration since there is no immediately apparent piece of data which falsifies either one. Both are simple, fairly plausible accounts of the ability to add. Either is falsified if it violates some known external constraint on the human neural system or if data exist that are incompatible with the theory. Whether either is anywhere near psychologically correct is outside the bounds of the present discussion. If we were interested in pursuing the question further, we would have to marshal a great deal of collateral evidence relevant to the following questions. Does one theory accord better with our set of intuitions about addition? Are there psychological experiments that might let us penetrate deeper? Does the time required to do a sum lend reasonableness to one theory but not the other? Further, have we been misled by external appearances to incorporate performance factors in what we intended to be a theory of competence?

Let us turn to linguistics, to which much of the preceding discussion is immediately relevant, though arithmetic ability and linguistic ability are crucially different in a number of essential and interesting ways. Our goal in

linguistics is the construction of a grammar: the correct account of the linguistic competence of the native speaker-hearer of a language. Note well that this goal is a matter of choice. Other choices of goals are possible and in fact have been made. It might be proposed that the goal of linguistic inquiry is to devise the best way of reducing languages to writing, though it is doubtful whether anyone ever has made any such outlandish suggestion. In principle there is no objection to a choice like this; one can only observe that it seems overly limiting and not very interesting.

We take then as our immediate goal in linguistics the formulation of a correct account of the speaker's intrinsic knowledge of his language—the system of rules that determine a connection of sound and meaning for each sentence in the language. This account is of the speaker's *competence*, not his *performance*. Whether he enunciates sloppily, whether he breaks off in mid-sentence and starts on a new topic, whether he (as a speaker of English) varies the amount of aspiration on his word-initial *p*'s—all these matters of performance are not within the accountability of the grammar.

This is not to imply that linguistics is unconcerned with questions of performance. Rather, we consider a performance theory as contingent on a competence theory, and consequently, since so much remains to be done even in the area of competence, we regard accounts of competence and performance as the immediate goal. The testability of a linguistic theory, however, requires that linguists and psycholinguists determine precisely what belongs to performance and what to competence. After all, it is through performance and the judgments associated with performance that we gain insight into competence, and it would not be surprising to find aspects of performance carried over by error or oversight into our accounts of competence. As with arithmetic ability, it is not always apparent where competence leaves off and performance begins.

The distinction between competence and performance, a necessary dichotomy in the investigation of any aspect of complex human behavior, has always existed in linguistics. It has in fact been observed almost universally, though only recently has it been so explicitly formulated and insisted upon. Saussure's famous distinction between *langue* and *parole* is partially analogous to competence and performance, though one need not accept those aspects of his *langue* (competence) that are essentially social, nor subscribe to Saussure's larger conception of language as a system of elements whose values are determined by all other elements in the system. In any case, when writing grammars, linguists have always stripped away at least the grossest of performance factors. No one has ever written parallel accounts of Hindi entitled *A Grammar of Hindi*; *A Grammar of Hindi As Spoken When Excited*; and so on.

Rather, linguists have always tried to describe what a speaker intuitively "knows," not how he uses that knowledge in this or that set of circumstances. This knowledge has universally been assumed to be present in some form or other, though not always consciously available to the speaker. A grammar—

i.e. a formal account of competence—should provide a starting point for formulating and testing theories of performance, but the grammar qua formal account of competence makes no direct commitment vis-à-vis performance.

We may now determine from our arithmetic example to what extent theory construction for arithmetic ability parallels construction of a grammar. What are the data of linguistics? They are in part quite similar to those involved in arithmetic: observations, usually in the form of phonetic transcription. But in linguistics there are other kinds of less tangible observations which are nonetheless relevant data. These are intuitive judgments of the native speaker, of which statements like the following are a tiny sample: "The sentences *John saw Mary* and *Harry eats apples* are not." "There is something peculiar about the plurals of English nouns like *foot, goose, child,* and *woman.*" "The sentence *Martin found the boy studying in the library* has several different interpretations." In other words, the native speaker of a language has acquired intuitions about his language that, along with his observable phonetic behavior, constitute an area of accountability of a theory of linguistic competence.

In common with accounts of arithmetic competence, a grammar must permit an infinity of results (sentences) with only finite means. For: (1) there is no upper bound to the number or length of sentences that a human being can in principle produce and understand (i.e. speakers can produce and hearers can understand new and novel sentences without any particular difficulty, and given most declarative sentences one can make a longer sentence by, for instance, preceding the sentence with *It is a fact that*), and (2) the brain is finite in capacity.

A grammar in the sense of "formal account of competence" must, given any particular sentence, state precisely and mechanically the steps involved in connecting meaning and sound. Such grammars are called *generative grammars*. Examples of grammars that are not explicit are abundant: for example, any school grammar of German, French, or Russian. Rather than describing the competence that underlies speech and linguistic judgments, pedagogical grammars generally give the paradigms and lexical items from which the reader can generalize to produce utterances not given in the paradigms and examples. Such a procedure is effective enough for learning languages, but it is obviously inadequate as an explicit account of the knowledge underlying the speaker-hearer's ability to use and produce an infinity of utterances.

The given substance for our work as linguists consists then of the primary data (roughly equivalent with phonetic observations about speech), linguistic judgments, and the external constraint that our theory must have a recursive property—i.e. must give explicit directions for producing an infinite set of utterances given only a finite number of rules and lexical entries (morphemes). Parts of all this have obvious analogues in our discussion of arithmetic competence, and we will now proceed to the framing of first accounts of

linguistic competence, but not before one crucial difference between arithmetic and linguistic investigation has been pointed out.

Arithmetic ability, because it is taught, can be consciously recreated to some extent: we can simply ask a subject how he does sums and he will tell us exactly what steps he goes through. This is not possible in linguistics. No linguist would get anywhere by asking someone exactly how he produces the sentence *John and Mary are going to get married.* Linguistic ability is acquired unconsciously as one part of the maturation process of a child, as is visual perception and crawling. We aren't taught such things the way arithmetic is taught and the knowledge underlying these processes is not available to our consciousness.

Notwithstanding this essential difference between arithmetic and linguistic ability, much of our account of arithmetic competence carries over easily to linguistics. We exclude performance data where possible and consider various plausible accounts of linguistic competence. An account that describes a finite corpus of primary data is said to meet the level of *observational adequacy*. A grammar that gives a correct account of the primary data *and* of the speaker's tacit knowledge—his intuitions and judgments about his language—is said to meet the level of *descriptive adequacy*. A descriptively adequate grammar, then, is both a physically correct account of the primary data and the psychologically correct account of the knowledge underlying these data and the linguistic judgments of the speaker-hearer. A linguistic theory (not a grammar) is said to meet the level of *explanatory adequacy* if it provides a principled basis for the selection of descriptively adequate grammars. Given any number of observationally adequate grammars, explanatory adequacy selects the descriptively adequate grammar.

To decide which of two observationally adequate grammars (or even grammar fragments) is closest in form and substance to the descriptively adequate account of the competence of the speaker is in practice extremely difficult, and in particular much more difficult than determining whether observational adequacy has been met. The latter problem is essentially a matter of goodness-of-fit between what the grammar predicts and the primary data, whereas the determination of descriptive adequacy is bound up with language acquisition and its numerous psychological factors. Hence no entirely explanatorily adequate theory is likely to be forthcoming in the near future. At any stage of psychological research and linguistic knowledge, we are not apt to know which grammar best recapitulates our true linguistic competence. Nevertheless, it is crucial to the development and refinement of linguistic theory to have in it an evaluation measure to enable us to pick out of any number of observationally adequate grammars the one that best attains the level of descriptive adequacy.

This conception of the goals of linguistic theory clearly moves linguistics into an intimate relationship with psychology, especially cognitive psychology that deals with the child's acquisition of language. Given a finite set of heard

utterances, a child develops a system of internalized rules which we call his linguistic competence. Similarly, under conditions such as those of the elementary classroom he develops a system of rules which we call his arithmetic competence. As linguists (or as students of arithmetic cognition) we can devise many observationally adequate accounts of the primary data. But which of these most closely resembles the one that has actually been constructed in his mind in the form of rules, elements, and so on? A theory approaches explanatory adequacy to the extent with which it offers a principled (i.e. not ad hoc) basis for making such a choice. Obviously we are very far removed from attaining explanatory adequacy on a large scale, and in fact we are very happy when we can bring it to bear even in a gross way on a given decision in linguistics; but this should not dim the insight that the relevance of linguistics to the other fields of human behavior grows in direct proportion to the degree of attainment of explanatory adequacy. And in linguistics proper our success in determining the descriptively adequate grammar of a language is directly proportional to the degree of explanatory adequacy of our linguistic theory.

Here and elsewhere in this book a systematic ambiguity is often used in speaking of a "grammar" (cf. Chomsky and Halle 1968:3–4). On the one hand, we mean the speaker's internally represented and organized body of intrinsic linguistic knowledge—his competence. On the other hand, we mean by "grammar" also the linguist's account of this competence—his formal, explicit, written account. The latter is a set of rules, elements, and so on, often written on paper; the former is immensely abstract and complex knowledge contained in the human organism. Until our evaluation measures are vastly more refined and explicit than they are now, there is no reason to suppose that there is point for point correspondence between the two.

In other words, the speaker has an internalized grammar—a competence—and the linguist's grammar is a *model* of this competence. The difference between these two uses of the word "grammar" will become important when we speak of linguistic change, for we will make statements of the sort, "Clearly what has happened here is that a rule has been added to the grammar of the speaker of Quechua." We do not mean that the speaker of Quechua has somehow consciously stuck a rule into his head thereby changing his (internalized) grammar. What we mean is that the primary data have become different between two types of Quechua, say two dialects or two chronologically distinct stages of the language; that is, the data—the phonological alternations in the languages, the positions of elements in the sentence, perhaps some judgment of the speaker—are different in a systematic way. *Our account of this difference is best described in terms of rule addition to the grammar. When we add the rule, our model of the competence of the Quechua speaker becomes current.*

The distinction of competence and performance is crucial not only to our notion of what a generative grammar is and is not, but in historical

linguistics is specifically crucial to our conception of linguistic change. Within generative grammar, change is regarded as change *in competence*, not just in performance. Change occurs because the grammar of the language has changed, and the largely random effects of performance have nothing to do with it. It is perhaps this empirically based conviction, rather than others which might be mentioned, that sets off the conception of historical linguistics within generative grammar from its conception within other linguistic traditions. This point will be argued and illustrated in Chapter 5. See also Postal (1968:271–281).

Finally, let us take brief cognizance of the role of linguistic universals in both general and historical linguistics. It is often observed that languages from all over the world share a surprisingly large number of common features. In phonology many five-vowel languages have the vowels /i u e o a/, but few if any have /i e ü ö æ/. The consonants /p t k s n/ are rarely absent in a language. Such properties of language, such "linguistic universals," determine the class of possible natural languages and the class of potential grammars for some particular language. We assume that these universal properties are in some way available to the child and that they are an integral part of the evaluation measure that selects for the child the best (descriptively adequate) grammar of his language. In this sense the search for linguistic universals is almost coterminous with the study of the innate ability that makes language acquisition in children possible in so short a time under conditions far from ideal (Chomsky and Halle 1968:4).

It is obvious to anyone acquainted with the state of current general linguistic theory that we are far from secure in our knowledge of any large body of linguistic universals. But their importance is not lessened nor should our concern for them be diminished. These assertions are equally valid for historical linguistics, where every aspect of work draws on our knowledge of what is universal in language—what is possible and likely in a natural language, what processes of change may be assumed as plausible.

Appeals to universals are most patent in linguistic reconstruction, where at every step toward the recovery of earlier structure we must ask ourselves whether the reconstructed stage is possible in terms of what we know to be generally true of natural languages—i.e. to be linguistic universals or near-universals. Any linguist who found himself reconstructing a proto-language with the two-vowel system /i e/ would do well to keep the news to himself unless the evidence for his reconstruction was uncommonly good. Why? Because the languages we know do not have vowel systems like this. Similarly, we would have a lot more faith in the reconstructed five-vowel system /i u e o a/ than /u i ʌ œ æ/, because the latter is more highly marked, more complex than the former (Chomsky and Halle 1968:409).

In another subfield of historical linguistics, that of genetic relationship, our judgments are obviously heavily influenced by knowledge of the universal properties of language. For example, no linguist is likely to be impressed by

the argument that two languages are genetically related because the word for 'father' in each begins with a labial obstruent such as [p]. Jakobson (1941) has shown that children, whatever their language, acquire sounds in a largely predictable order ([p] being one of the first) dependent not on external factors like frequency of exposure but only on the intrinsic complexity of sounds. For universal reasons, therefore, totally unrelated languages may have words beginning with [p] or [m] to designate those beings on whom a child is initially most dependent.

Our work as historical linguists is narrowly constrained by our judgments of what is and what is not a universal property of natural language, and we can expect the progress of historical linguistics to be closely connected with the search for linguistic universals.

2.2 THE FORM OF A GRAMMAR

Following Chomsky (1965:15–18) and Chomsky and Halle (1968:6–7), we shall assume that the grammar of any language must have the following major components:

(1) a syntactic component, consisting of
    (a) a base component, and
    (b) a transformational component;

(2) a semantic component;

(3) a phonological component.

The position of these components in the grammar may be represented schematically as in Figure 2.1.

The *base rules* and the *lexicon* make up the *base component* of the grammar. Various other rules, such as the "readjustment rules" which relate syntax to phonology (Chomsky and Halle 1968:10), are not discussed here.

THE BASE RULES. This system of rules generates the restricted set of basic sentence types in the language together with the structural description that is associated with each basic sentence type. The base rules thus correspond in part to the phrase-structure rules or constituent-structure rules of earlier formulations of transformational grammar (Bach 1964:33–53 and Chomsky 1957:26–33). The structural description may be represented in part by a labeled branching diagram or phrase-marker, as in Figure 2.2.

The output of the base component is the deep structure of the sentence. Figure 2.2 represents a portion of the deep structure of the sentence whose ultimate phonetic representation would be approximately [aɪ tʰoʊl jæk tʰə kʰʌm] 'I told Jack to come'. A number of operations would have to be performed on the deep structure of Figure 2.2 to obtain from it this
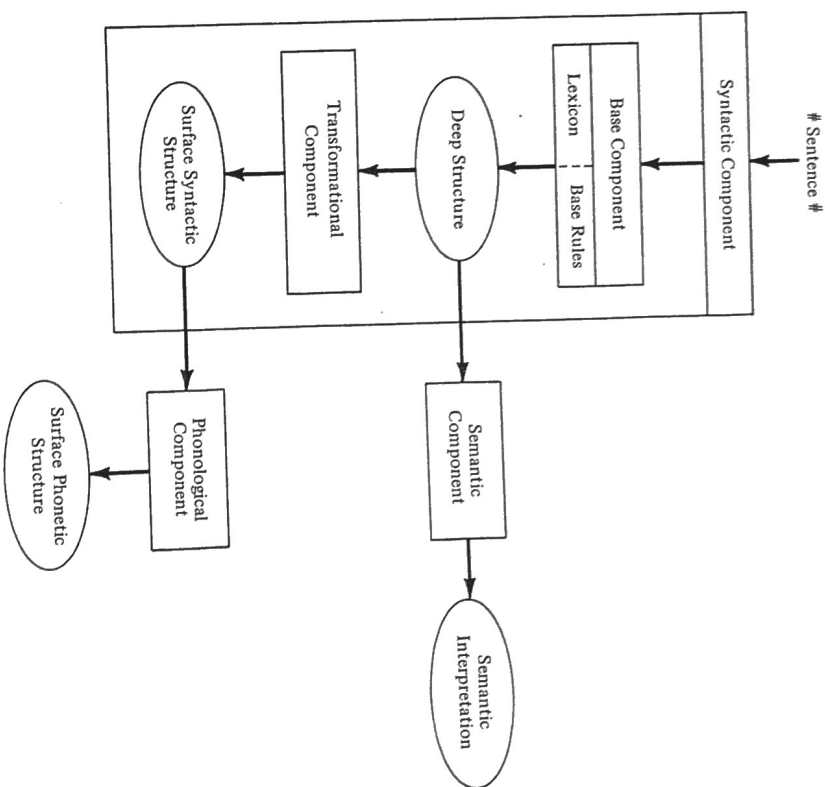
FIGURE 2.1
ORGANIZATION OF A TRANSFORMATIONAL GENERATIVE GRAMMAR

phonetic surface structure. Transformational rules would delete the second occurrence of *Jack*, convert Pres to *to*, and carry out various other rearrangements in the string produced by the base component. The phonological component contains rules that specify the final phonetic form of this utterance.

THE LEXICON. This component of the base is identifiable in part with the traditional notion of dictionary. It consists of a list of the morphemes of the language and the information that characterizes the behavior of each morpheme at all levels. A partial statement of such information for each morpheme would be the redundancy-free underlying phonological shape expressed as a matrix of distinctive features, its syntactic category, information required for the semantic interpretation of the sentence, any individual peculiarities that

(Figure 2.2: tree diagram)

Node labels, top to bottom:
S
— NP (I)
— Aux (Past)
— VP
  — V (tell)
  — NP (Jack)
  — S
    — NP (Jack)
    — Aux (Pres)
    — VP
      — V (come)

FIGURE 2.2
PHRASE-MARKER FOR *I told Jack to come.*
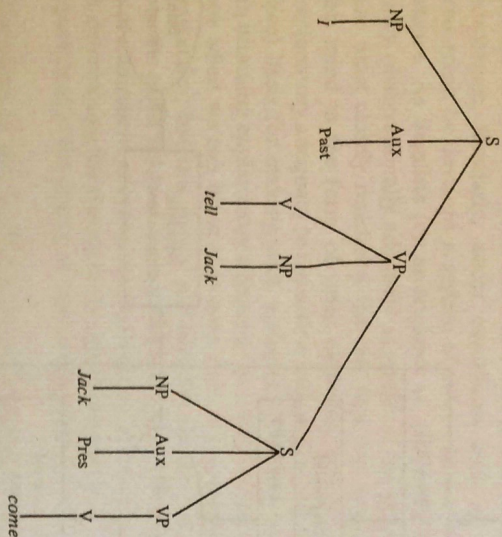
(The abbreviations used here are as follows: S = Sentence, NP = Noun Phrase, Aux = Auxiliary, VP = Verb Phrase, V = Verb. The italicized items such as *I, tell,* and *Jack* are taken from the lexicon.)

make the morpheme violate constraints normally placed on morphemes of its general type. To illustrate a typical entry in the lexicon, and at the same time to emphasize the abstract nature of the phonological representation of morphemes, consider the deep structure for the sentence *Divinity may frighten the boy* (cf. Chomsky 1965:108–109 for a similar example), shown in Figure 2.3.

Entries such as *divinity, may, frighten,* and *boy* in this deep structure come from the lexicon. The item *divinity* has a lexical representation that the following approximates:

(divin+i+ty   [+ Noun, + Determiner ———   – Count,
                                            + Abstract,...])

That is, the noun *divinity*, whose underlying phonological representation is representable informally as /divin+i+ty/ where + stands for "morpheme boundary," is not a Count Noun but an Abstract Noun, and so on. The three dots ... cover the host of other properties that characterize the behavior, syntactic, semantic, and phonological, of the particular noun *divinity*.

(Figure 2.3: tree diagram)

Node labels:
S
— NP
  — Determiner
    — Article
      — [–Definite]
  — N (divinity)
— Predicate Phrase
  — Aux
    — Modal (may)
  — VP
    — V (frighten)
    — NP
      — Det
        — Article
          — [+ Definite]
      — N (boy)
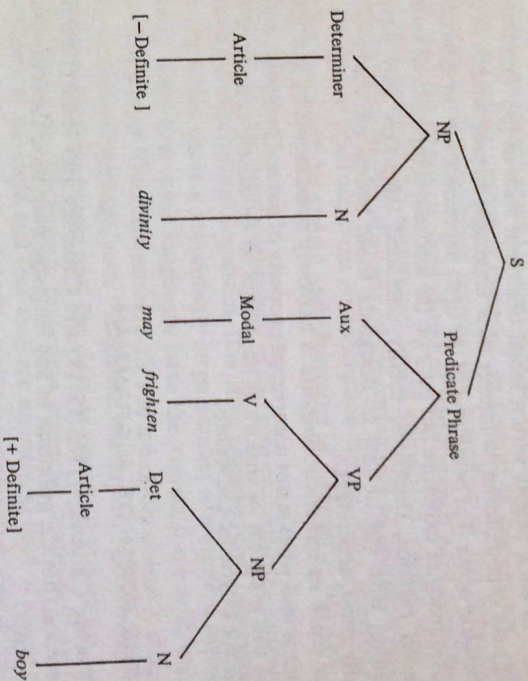
FIGURE 2.3
PHRASE-MARKER FOR *Divinity may frighten the boy.*

The spelling /divin+i+ty/ is, as mentioned, only an informal representation of the phonological shape. Its true representation in the lexicon is a sequence of bundles of distinctive features—the minimum set of distinctive feature specifications needed to account for the final phonetic shape of this word, namely [dívìnìty]. Thus, *divinity* appears in the lexicon in a representation that Figure 2.4 approximates.

The distinctive features used in specifying these segments and boundaries are taken from Chomsky and Halle (1968). Further discussion of distinctive features will follow under the rubric of the phonological component.

Associated with the lexicon is a set of *contingency statements*, which fill out redundant feature specifications in the lexical representations of morphemes (cf. Stanley 1967). These statements, which correspond to the rules of the morpheme-structure component of earlier formulations of generative phonology (Harms 1968), specify two kinds of redundancy. First, they specify segmental redundancies of the sort embodied in statements such as: all vowels, nasals, and resonants in English receive the specification [+ voice]; vowels specified as [+ back] in English are redundantly [+ round]. Second, they fill out the redundancies that result from sequential constraints on the occurrences of segments. In English, if the first C in the morpheme-initial sequence +CC is an obstruent, then the second C must be [s] (as in *spin*); if the second C is a liquid, then the first C must be an obstruent (*play, crime*).

FIGURE 2.4

PHONOLOGICAL REPRESENTATION OF *divinity* IN THE LEXICON

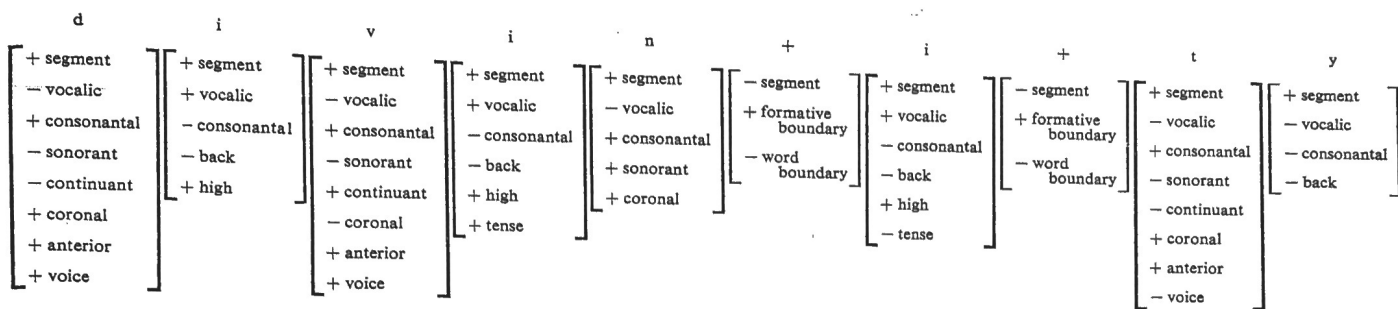The phonological representation of morphemes in the lexicon is thus extremely abbreviated and quite abstract. Redundancy statements add features to the redundancy-free representations and produce fully specified distinctive feature matrices which then are the input to the phonological component of the grammar.

It should be noted that the structure of the lexicon is currently subject to much debate. For example, for a variety of reasons phonological representations in the lexicon should be stated in terms of *marked* and *unmarked* values of features rather than strictly in terms of pluses and minuses, as was done above in accordance with phonological practice current through the mid 1960's. For a preliminary discussion of this question see Chomsky and Halle 1968:400-435).

THE SEMANTIC COMPONENT. This component assigns semantic interpretations to the deep structure generated by the base component. These interpretations, called "readings," correspond to the traditional notion of "possible meanings of an utterance." Although semantic change is of great intrinsic interest in historical linguistics, this book has nothing to say about it. For discussion of the semantic component of a grammar see Katz and Fodor (1963) and Katz and Postal (1964).

THE TRANSFORMATIONAL COMPONENT. The rules in this part of the grammar convert the deep structure generated by the base component into the surface syntactic structure. Some transformations have been alluded to in the discussion of the base. In general, transformations effect changes in the order of elements in the underlying string produced by the base, insert elements into the string, delete others—in short, all changes required in producing the surface syntactic structure from the deep structure of a sentence. Literature on the transformational component is now voluminous, cf. Bach (1964), Katz and Postal (1964), and Chomsky (1957, 1965), and acquaintance with its operation is assumed.

THE PHONOLOGICAL COMPONENT. The rules in this component of the grammar act on the surface-structure syntactic string produced by the syntactic component and provide a phonetic interpretation for the sentence represented by this string. Let us consider again the example *Divinity may frighten the boy* (cf. Figure 2.3).

Rules in the transformational component are required to specify the shape of [− Definite] in various contexts. [− Definite] is realized as null before the [− Count] noun *divinity*, and [+ Definite] is realized as *the*, phonologically /ði/, before *boy*, which is [+ Count]. When these and any other required transformations have been carried out, the string which leaves the syntactic component as input to the phonological component will have a structure as in Figure 2.5:
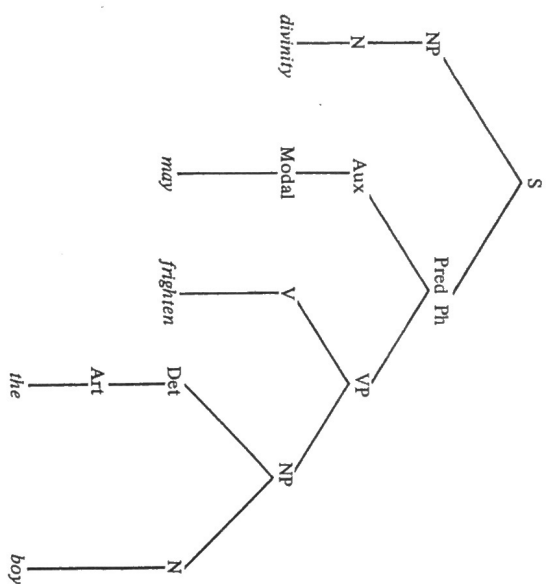
or equivalently as:

[S[NP[N+*divinity*+]N]NP[Pred-Ph[Aux[Modal+*may*+]Modal]Aux[VP[V+
frighten+]V[NP[Det[Art+*the*+]Art]Det[N+*boy*+]N]NP]VP]Pred-Ph]S

FIGURE 2.5

SURFACE SYNTACTIC STRUCTURE OF
Divinity may frighten the boy.

```
                          S
            ┌─────────────┴─────────────┐
           NP                        Pred Ph
            │                 ┌──────────┴──────────┐
            N                Aux                    VP
            │                 │              ┌───────┴───────┐
        divinity           Modal            V              NP
                             │               │        ┌──────┴──────┐
                            may          frighten    Det            N
                                                      │             │
                                                     Art           boy
                                                      │
                                                     the
```

In addition to the syntactic information retained in this representation, any data relevant to a single morpheme are carried over from the lexicon. If, for example, some lexical item in this string is an exception to a particular phonological rule (perhaps because that item has recently been borrowed from a foreign source, or for any reason whatsoever), then this information is retained in the string entering the phonological component and treated as a signal that the aberrant morpheme is not to undergo the rule that would normally apply.

In general, syntactic information is necessary for the operation of at least certain phonological rules. English, as well as German and various other languages, has a set of rules that assign stress. These rules operate cyclically, in "the transformational cycle," and syntactic bracketing is required for their operation. On the transformational cycle see Chomsky and Halle (1968:59–162) and Harms (1968). Rules that require syntactic information are applied as the first set of rules

in the phonological component. As they are applied, the syntactic bracketing is "erased" from the underlying representation. After the last phonological rules in the cycle have been applied, syntactic information is not available for the phonological rules which follow. At this point in the derivation of the sentence *Divinity may frighten the boy*, we are left with a string of phonological elements which may be represented as

$$\# \#\text{divīn}+\text{i}+\text{ty} \# \#\text{mæ} \# \#\text{frit}+\text{n} \# \#\text{ði} \# \#\text{bǣ} \# \#$$

The symbol # represents *word boundary*, the symbol + represents *morpheme boundary*, and formatives such as *divinity* and *may* are presented in an orthography intended to convey their underlying phonological representations. For explanation of the not immediately obvious aspects of these phonological base forms, e.g. /bæ/ for surface phonetic [boy], see Chomsky and Halle (1968).

The remaining phonological rules may be divided into two blocks: (1) the phonological rules proper, which convert the underlying systematic phonological representation into a "low-level" phonetic representation, and (2) phonetic rules that assign detailed phonetic specifications to the previously obtained string of phonological segments. The rules in the first block of rules are assumed to operate on and produce segments stated in terms of *binarily specified* (either + or − , but no other value) distinctive features. The second block of rules produces segments whose intersecting set of distinctive features may contain *n-ary* values of these features: not necessarily only + and − , but values specified in terms of a sequence of integers.

Returning to the underlying phonological representation of *Divinity may frighten the boy*, phonological rules will carry out many kinds of changes needed to obtain the surface phonetic structure. These rules are discussed and motivated in depth in Chomsky and Halle (1968), and here they will be only indicated in brief. Phonological rules would diphthongize /ī/ > [ay], as in /frīt+n/ > [fraytn], another rule would lax /i/ > [I] under certain conditions, as in /divin+i+ty/ > [divinity], and so on. (Here a condition has been introduced and will be followed concerning the use of solidi // and square brackets []. Solidi enclose systematic phonological segments—the informally designated symbols used to spell items in the lexicon. Square brackets enclose segments that have been produced by rules in the phonological component of the grammar.)

We will state phonological rules in terms of distinctive features, but for convenience each rule will be followed by a plain English statement of what the rule does. The present book depends on a particular set of phonological features only in a rather peripheral way. Mostly we will be concerned with rules and various sorts of rule phenomena. It is understood that the basic units manipulated by these rules are distinctive features, but the use of any particular set of features is not crucial. For this reason, the features used

throughout this book are (with one or two exceptions) chosen to minimize the difficulty for the reader who has not undergone and mastered the vicissitudes of shifting feature terminology during the 1960's. For vowels, for example, the familiar features *high, low, back, round, tense,* and *long* are used. For consonants the features used in the majority of rules should not create problems for anyone acquainted with phonetic terminology: e.g. *nasal, continuant, lateral, palatalized,* and *voiced.* The features *obstruent, vocalic,* and *consonantal* are used to define the major natural classes of segments as follows:

	vowels	obstruents	nasals	resonants	glides
obstruent:	-	+	-	-	-
vocalic:	+	-	-	+	-
consonantal:	-	+	+	+	-

Thus, the class of vowels is defined by $\begin{bmatrix} + \text{vocalic} \\ - \text{consonantal} \end{bmatrix}$; the class of obstruents (true consonants such as *p d k s j*) is defined by [+ obstruent]; *r* and *l* are $\begin{bmatrix} + \text{vocalic} \\ + \text{consonantal} \end{bmatrix}$; and glides such as *j* and *w* are $\begin{bmatrix} - \text{vocalic} \\ - \text{consonantal} \end{bmatrix}$.

The features *anterior* and *coronal* (adopted from Chomsky and Halle 1968) for consonants require some explanation. *Anterior,* which is identical with *diffuse* for consonants in earlier distinctive feature theory, describes sounds produced with an obstruction located in front of the palato-alveolar region of the mouth. Thus, the following classes of sounds are [+ anterior]: labials (*p b m*), alveolar and dental consonants (*t d n*); and the following are [- anterior]: palato-alveolars (*č j ñ*), velars and uvulars (*k g ŋ x*). Coronal sounds are produced with the blade of the tongue raised from the neutral position. Labials, velars, palatals, and uvulars are [- coronal], whereas dentals and palato-alveolars are [+ coronal]. Dental [r] is [+ coronal], and uvular [R] is [- coronal].

Often, for the sake of convenience, we will write V and C to denote respectively the natural classes of "vowels" and "consonants." Thus, the segment

$$\begin{bmatrix} V \\ + \text{high} \\ + \text{round} \\ + \text{back} \\ - \text{long} \end{bmatrix}$$

designates a vowel which is high, round, back, and short—i.e. [u]. The

$$\begin{bmatrix} C \\ + \text{anterior} \\ - \text{coronal} \\ + \text{nasal} \end{bmatrix} \quad \begin{bmatrix} C \\ + \text{anterior} \\ + \text{coronal} \\ - \text{continuant} \\ + \text{voice} \end{bmatrix} \quad \begin{bmatrix} C \\ - \text{anterior} \\ - \text{coronal} \\ + \text{continuant} \\ - \text{voice} \end{bmatrix}$$

could refer (depending on the language and what segments occur in the language) respectively to [m], [d], and [x].

The segments or natural classes stated so far contain distinctive features marked either plus or minus. The n-ary rules, the rules that assign detailed phonetic specification to segments, will assign more than binary specifications of features where needed. In English, for example, both [a] and [ɔ] are [+ back], but [ɔ] is typically more back than [a]. Hence, a rule in the last block of rules in the phonological component—the set of rules that assign final, detailed phonetic specifications—will assign [ɔ] a higher backness value than [a]. The conventions for such assignments are not well established, but the principle involved may be indicated by assuming that a phonetic rule will assign the value [+2 back] to [ɔ] and [+1 back] to [a].

For a detailed discussion of the phonological component of a grammar see Harms (1968) and Chomsky and Halle (1968). Putting aside questions of the formal appearance of the phonological component, we may say that its content represents a set of generalizations about the phonological structure of a language. These generalizations are stated in the form of rules that act on underlying abstract representations of sentences to produce a phonetic interpretation.

The chief source of evidence for deciding which rules to posit is provided by phonological alternation—or, as it is often called, morphophonemic alternation—in the language. The presence of phonological alternation is usually an indication that a generalization can be extracted by positing a common underlying form and positing a rule or set of rules that accounts for the phonetic realizations of the members of the alternation. In German, voiceless obstruents word-finally alternate with voiced obstruents in non-word-final position: *Rad* [ra:tʰ] 'wheel' versus *Tage* [tʰa:gə] 'days', *Gras* [gra:s] 'grass' versus *Rades* [ra:des] 'of the wheel', *Tag* [tʰa:kʰ] 'day' versus *Rades* [ra:des] 'of the grass'. This completely general phonological alternation supports the decision to derive both forms participating in such alternations from common phonemic sources; that is, both the [tʰ] in *Rad* and the [d] in *Rades* should be derived by rule from a single phonological segment. Whether this segment should be /d/, /t/, or a segment different from

either must be determined by simplicity. In the German case, we take the voiced member of the alternation as basic and derive the voiceless member by means of the rule:

$$[+ \text{obstruent}] \rightarrow [- \text{voice}] / \_\_\_ \#$$

(Obstruents are devoiced in word-final position.) This solution provides derivations like the following:

Underlying Form:	/rad	rades	tag	tage	gra:z	grazes/
Devoicing:	rat	...	tak	...	gras	...
Phonetic Form:	[ra:t	ra:dəs	tʰa:kʰ	tʰa:gə	gra:s	gra:zəs]

(This partial derivation is presented only to show how the voiceless/voiced alternation arises. Other phonological rules would be required to account for the lengthening of the vowel, for the occurrence of [ə], and for the aspiration occurring in various positions.)

It is crucially important to understand that the devoicing rule is included in the synchronic grammar of German because of certain facts about German today, not because terminal devoicing occurred in the historical development of German. Given the voiced/voiceless alternations found in modern standard German, the simplest way of accounting for them is to include a rule of terminal devoicing in the grammar. We may call this a "late" rule meaning that it need not necessarily precede certain rules in German and must in fact be placed close to the end of the set of ordered rules comprising the phonological component. It just so happens that the devoicing rule entered the grammar of German a long time ago—slightly less than a thousand years. Chronologically, this makes devoicing a rather early rule compared with others found in the grammar of modern German, but this does not alter our judgment that terminal devoicing is a rule in the synchronic sense and a late rule at that.

2.3 POSTSCRIPT

It should be emphasized that what has been said here about the internal make-up of a transformational generative grammar is tentative in many of the details. Like any serious effort to describe complex modes of human behavior, the theory of generative grammar is not static and unchanging. New perceptions, new data, and old data seen in a new light lead to revisions and corrections in one or the other component. The new predictions, general and particular, created by the new grammar and the theory may in turn bring new data to light, and these data may again suggest new revisions in the grammar and the theory.

This is a natural process in the development of any science. It is natural, it is unavoidable, it is not cause for grief. The general picture presented here of the form of transformational generative grammar is adequate in most points, so far as we know at present, though a considerable lack of clarity exists in details. The line of demarcation between the syntactic and the semantic components, the form of the lexicon, the set of distinctive features—these raise questions as yet unanswered. But despite these uncertainties, this chapter's theory of language and grammar provides much insight into language and, as well, into historical linguistics, to which we now turn.

SUPPLEMENTARY READING

Bach, Emmon. 1964. An Introduction to Transformational Grammars. New York: Holt, Rinehart & Winston, Inc.

Chomsky, Noam. 1964. "Current Issues in Linguistic Theory," in The Structure of Language, eds. Jerry A. Fodor and Jerrold J. Katz. Englewood Cliffs, N.J.: Prentice-Hall, Inc.

Chomsky, Noam. 1965. Aspects of the Theory of Syntax. Cambridge, Mass.: The M.I.T. Press.

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

Harms, Robert T. 1968. Introduction to Phonological Theory. Englewood Cliffs, N.J.: Prentice-Hall, Inc.

Jacobs, Roderick A., and Peter S. Rosenbaum. 1968. English Transformational Grammar. Waltham, Mass.: Blaisdell Publishing Company.

Katz, Jerrold J., and Paul M. Postal. 1964. An Integrated Theory of Linguistic Descriptions. Cambridge, Mass.: The M.I.T. Press.

Langacker, Ronald W. 1968. Language and Its Structure. New York: Harcourt, Brace & World, Inc.

Postal, Paul M. 1968. Aspects of Phonological Theory. New York: Harper & Row, Publishers.

3

PRIMARY CHANGE

The topic of central interest in historical linguistics is linguistic change. If languages never changed, clearly no one would be interested in the historical development of languages except in connection with political and social events. Whatever else we may do in the name of historical linguistics, at bottom we are dealing in matters at first or second remove from change in a language or language family. But before saying what we can about linguistic change, we shall make what at first glance seems a detour to concern ourselves with the question of dialects and their differences. This is really not a digression because any change—our paramount concern—is ultimately rooted in the process of two dialects having become different. Dialects, in other words, provide the most direct evidence regarding change at our disposal. Thus, our immediate job is to clarify the notion of difference between related dialects of the same language, for without understanding dialect differences we

cannot hope to determine with any precision how those differences have arisen; that is to say, we cannot determine what has changed or indeed for that matter what change really is.

3.1 ON THE DIFFERENTIATION OF DIALECTS

The problem of describing the differences between dialects, especially at the phonological level, was given much attention by linguists during the 1950's, and a great deal of agonizing went on in its behalf. The crux of the dilemma was that dialects seemed to be essentially nondichotomous in structure. They blend almost imperceptibly into one another within a given language area. In fact, dialects seem to constitute a perfect example of the kind of continuity that, as Martin Joos remarked, is "shoved outside of linguistics in one direction or another" (1950:702). Typically, structural linguistics dealt with this sort of continuity by quantizing—by assigning continuous data to a discrete unit such as the autonomous phoneme. (Cf. Postal 1968 on the "autonomous" phoneme.) Thus, the autonomous phoneme /a/ in English was regarded as a discrete abstract unit realized in countless ways all clustering around a phonetic norm [a]. The data on performances of /a/ make up a continuum of observations, yet as structural linguists we would have focused our concern first on procedures which allow us to deal with a finite, and fairly small, number of discrete units such as autonomous phonemes that have a functional role in language. The continuity would be left to phoneticians.

Quantization of this kind is a characteristic and crucial part of American descriptive ("Neo-Bloomfieldian") linguistics, and it is precisely this sort of dichotomous procedure that dialects seem to be the least amenable to. That is to say, dialects seemed inherently beyond the procedures of Saussurean and Bloomfieldian linguistics, both of which assumed that typical units of structural description (phonemes, morphemes) were defined *within an idiolect* by their relationships to all other units in the system of that idiolect.

Let us consider an example. Within autonomous phonemics, the vowel /a/ is a phoneme in my idiolect because it contrasts with the vowels of *pet, pat, put*, and so on. Another speaker of American English may have an autonomous phoneme /a/ that likewise is defined by contrast with other units in his speech. In what sense can one "compare" /a/ in the two idiolects? In what sense is it legitimate to assert that this other speaker and I have the "same" phoneme /a/ in our speech? The only allowable answer in strict descriptive linguistics is "None" unless we have speech patterns that are identical in every single respect, a condition that has never been observed in the speech of any two people anywhere. For the autonomous phoneme /a/ is determined not by the fact that it generally sounds like [a] and thus is similar in realization to /a/ in other idiolects, but by the units it contrasts with in the whole of my speech.

Since the whole of my speech is different, though perhaps only in minute details, from that of anyone else, it necessarily follows that nobody else can possibly have a comparable phoneme /a/. The situation is made more apparent if we assume that I have no barred-i phoneme /ɨ/ in my autonomous phonemic inventory whereas the other speaker does. In this case my /a/ is clearly different from his since mine does not have an /ɨ/ to contrast with, and there is no basis for comparison of the "same" sound /a/ in the two idiolects.

In other words, how can we speak of dialect differences and dialect similarities? If we are strict about it, there seems to be no way of incorporating the "emic" units of descriptive linguistics into the description of dialect differences without violating the fundamental proposition of Saussure (which is implicit, though not always explicitly stated, in all schools of descriptive linguistics) that structural units are defined within an idiolect by the web of relationships locking them into the system.

From this strict point of view a number of things desirable in linguistics become impossible. It is impossible to compare two dialects, or two idiolects for that matter. It doesn't even make much sense to say that two dialects are different, since by definition a linguistic description is good for one, and for only one, speaker. And since what is true of dialect differences applies *mutatis mutandis* to differences between chronologically different stages of the same language, then how can we compare the differences between two stages of a language, say between Middle English and Modern English?

If we relax our observance of the Saussurean dictum (though exactly how this can be done is not immediately apparent), then the methods of structural linguistics (and specifically of autonomous phonemics) can be applied to dialects, but this always requires compromise with Saussure's fundamental principle. Several attempts at dissolving the inherent antinomy between dialects and structuralism emerged from the dialogue of the 1950's, and we shall discuss briefly two of these attempts here.

One solution widely subscribed to for a time was the "overall pattern." This was constituted by the minimal set of autonomous phonemes that taken together would account for all of the contrasts found in the speech of any speaker in a given language area. Each dialect, and by implication each idiolect, would then choose some subset of the maximal set of phonemes in the overall pattern.

This way of reconciling structuralism and dialectology has numerous bad consequences, which need not be enumerated here. The theoretical objections are obvious: besides violating the cardinal tenet that each dialect must be analyzed in its own terms without reference to or data from other dialects, what sense does it make to say that someone has in the overall pattern of his language a phoneme that he does not use? Merely because someone in some dialect of American English has phonemic /ɨ/—say on the basis of the usual

minimal pair [jɪst] 'just (adverb)' ≠ [jʌst] 'just (adjective)'—must I accept barred-i in the description of my speech and transcribe, say, [bɪznɪs] 'business' as /bɨznɨs/, even though for me [ɨ] is an allophone of some other vowel under weak stress and not phonemic? This seems radically counter-intuitive to the notion of "phoneme."

Practically speaking, too, the notion of overall pattern leads to conclusions that seem out of line with our intuitions about language and its structure. Consider the vowel nucleus in a word like *ride*. For a Wisconsinite this might be pronounced [aɪ], for the Mississippian [a<ə]. In the Trager-Smith (1957) overall pattern we would transcribe these /ay/ and, with a slight modification of the Trager-Smith system, /Ah/, where /A/ stands for a vowel approximately midway between /æ/ and /a/. Thus, the Wisconsinite has phonemic /rayd/ and the Mississippian /rAhd/. The overall pattern transcription shows that the two dialects pronounce the diphthong differently. But what is not stated is that both the Wisconsinite and the Mississippian have in their phonemic inventories only a single diphthong /ay/ contrasting with other diphthongs /aw/ and /oy/, so writing /ay/ for one dialect and /Ah/ for another implies a greater phonemic difference in the structure of the dialects than seems warranted by the data. Indeed, the two dialects have very similar phonemic inventories, and the main differences between them are at the phonetic level. That is, the phonemicizations /ay/ and /Ah/ emphasize dialect difference at the expense of the considerable dialect similarity. One could better say that both dialects have phonemic /ay/ which is realized by the Wisconsinite as [aɪ] and the Mississippian as [a<ə]. But this shoves whatever systematic pattern exists down into the phonetics of the description, and the eminently *structural* level of description—the phonemic level—fades in relevance to the characterization of the differences between these dialects.

A somewhat different approach to resolving the conflict between dialectology and structural linguistics was proposed in the article "Is a Structural Dialectology Possible?" by Uriel Weinreich (1954). (To be precise, we should say that Weinreich's goal was not so much to reconcile the two as it was to devise ways of accommodating the interests of dialectologists *and* linguists within the field of general linguistics.) His principal contribution was the concept of "diasystem." This is similar to, and in some senses a generalization of, the concept of overall pattern, but with the crucial difference that it characterizes dialect differences by the use of phonemic correspondences and not by the idea that dialects choose from among a set of abstract elements.

To illustrate the idea of diasystem, let us return to our *ride* example. We assume that, except for the difference in pronunciation of the diphthong in words such as *ride*, the Wisconsin and Mississippi dialects are identical; e.g. for *bit* both have [bɪt]. Thus, the two dialects share (in some sense or other which is not specified exactly) the phonemes /b/, /ɪ/, /t/, and all the others, but [aɪ] in Wisconsin corresponds to [a<ə] in Mississippi. There is only one /ay/

phoneme in either dialect, but its realization is different. The diasystem abstracted from this would be

$$\|b \approx \imath \approx \imath \approx t \approx \ldots \approx \frac{w[a\underline{i}]}{m[a\,\sp{'}\underline{\partial}]}\|$$

In this way we obtain a comparison between the two dialects that displays their points of agreement *and* their points of dissimilarity; and this is done without obscuring either the phonetics or the phonemics of the situation. Several objections, however, have been raised against the diasystem. It does not, nor did Weinreich claim that it did, get around the Saussurean riddle. But even if we ignore this forbidding crisis of theory, serious problems in application arise. The chief question is whether we take account of cognate items in the two dialects when we determine the diasystem. Ignoring cognates altogether leads to the absurdity that any two languages with identical phoneme inventories share the same diasystem. For example, Spanish and Standard Yiddish could be regarded as sharing the identical (phonemic) vowel diasystem $\|i \approx e \approx a \approx o \approx u\|$. This is obviously an undesirable result since the two are not dialects of the same language. On the other hand, requiring the two dialects to have their variants of the same diaphoneme in cognate items rules out the possibility of setting up a diasystem for different languages like Spanish and Yiddish, but again we are led to counter-intuitive results. This has been demonstrated by Moulton (1960:176-177). By imposing this condition he showed that two Swiss German dialects not fifty miles apart and mutually completely intelligible have no more than three shared diaphonemes (the dialects separately have eleven phonemes each), only one of which is fully shared.

This excursus into the problems of structural dialectology was made in order to point up the fact that it is by no means obvious how structural linguistics (or any theory of linguistics) is relevant to the description of the differentiation of dialects. Our real problem is how to account within a single linguistic theory for the essential fact about dialects—that they are in many ways similar—without unduly emphasizing the undeniable fact that they are in some ways different. The task is not an easy one, as the weaknesses of the overall pattern and the diasystem illustrate.

3.2 DIALECT DIFFERENTIATION IN GENERATIVE GRAMMAR

We shall now examine the implications that the goals and form of generative grammatical theory have for the description of dialect differences—a topic that, let it again be emphasized, is as pertinent to historical linguistics as it is to dialectology.

For the time being we shall confine our attention to phonology. We saw in Chapter 2 that our paramount concern as linguists is with the grammar of a

language—the system of rules that account for the native speaker's intuitions about his language and for the primary data. In particular, as regards the phonological component of a grammar, our main questions are (1) what is the set of underlying systematic phonemes?, and (2) what is the set of rules that capture with the greatest generality and economy the observable facts about the phonology of a given language? We saw also that the most direct way of getting at these rules was through the evidence of phonological (morphophonemic) alternation.

Let us now consider a hypothetical example in which two dialects of a language differ in a very minor way. In Dialect A there is an underlying set of five (systematic phonemic) vowels /i e a o u/. The only alternation of any interest at the present is $a : æ$. [æ] occurs after palatalized consonants, [a] elsewhere, e.g. [ap] 'dog', where $t_,$—is a prefix for indirect objects ($t_,$ denotes a palatalized t). We classify the five underlying segments as follows, where redundant features are enclosed in parentheses:

	i	e	a	o	u
consonantal:	–	–	–	–	–
vocalic:	+	+	+	+	+
high:	+	–	–	–	+
back:	–	–	+	+	+
low:	(–)	(–)	(+)	(–)	(–)
round:	(–)	(–)	(–)	(+)	(+)

To account for the $a : æ$ alternation we have the rule:

3.1
$$\begin{bmatrix} - \text{consonantal} \\ + \text{vocalic} \\ + \text{low} \end{bmatrix} \rightarrow [- \text{back}] / [+ \text{palatalized}] \text{---}$$

(A low vowel is fronted following a segment which is palatalized.) Next let us assume that Dialect B of this hypothetical language is identical in every respect with Dialect A except that forms having the $a : æ$ alternation in Dialect A have an alternation $o : æ$ in Dialect B before labial consonants, e.g. [op] 'dog': [t,æp] 'to the dog'. Note that Rule 3.1 affects only the low vowel /a/ in Dialect A. It does not apply to /o/, which is [– low], so that /o/ undergoes no fronting after palatalized consonants in Dialect A, e.g. [op] 'cat': [t,op] 'to the cat'. Similarly, in Dialect B words containing /o/ corresponding to /o/'s in Dialect A do not enter into any alternation after palatalized consonants; only those pre-labial [o]'s corresponding to [a]'s in Dialect A do this. Thus, Dialect B also has [op] 'cat': [t,op] 'to the cat'.

How shall we describe the situation in Dialect B? The only real trouble is that some [o]'s alternate with [æ]'s while others do not. It would not do to write 'dog' /op/ in Dialect B and assume operation of Rule 3.1 since this would not yield the correct form [t,æp] 'to the dog'; Rule 3.1 does not affect /o/, so we would get the wrong form *[t,op]. One conceivable solution would be to have two different kinds of /o/ segments in the underlying forms: /o/'s which do not alternate will be a plain /o/, those which do will be marked with an apostrophe, /o'/. To distinguish the two, as we must do in order to make them distinct in lexical morphemes, we will posit a feature "alternating" and classify the segments in Dialect B as follows:

	i	e	o'	o	u
consonantal:	−	−	−	−	−
vocalic:	+	+	+	+	+
high:	+	−	−	−	+
back:	−	−	+	+	+
alternating:	(−)	(−)	(+)	(−)	(+)
round:	(−)	(−)	(+)	(+)	(+)

We must now have a more complicated version of Rule 3.1, namely:

$$3.1' \quad \begin{bmatrix} -\text{consonantal} \\ +\text{vocalic} \\ +\text{alternating} \end{bmatrix} \rightarrow \begin{bmatrix} -\text{back} \\ -\text{round} \\ +\text{low} \end{bmatrix} / \,[+\text{palatalized}]\,\text{—}$$

(/o'/ becomes [æ] after palatalized segments.)

This solution does seem to account for the intrinsic knowledge of the native speaker of Dialect B that words containing certain kinds of /o/ sounds have grammatically related forms with [æ] when a palatalized consonant precedes. What is wrong with this solution? First, the feature "alternating" is strictly ad hoc, invented for this problem. It is used only to set off one kind of /o/ from another, and it has no discernible physical manifestation since there is no phonetic difference in the two /o/'s. This gives us a completely abstract phonemic representation divorced from the phonetic substance of which phonological alternations consist. The feature "alternating" imposes no natural constraints on phonemic representations like those implicitly imposed by our use of the features "high," "continuant," "back," and so on. A second, related objection to Rule 3.1' is that it disguises the basic phonetic naturalness of the process taking place in Dialect B. Rule 3.1 in Dialect A expresses a natural process frequently encountered in the languages of the world: a vowel is fronted after a palatalized consonant (Russian has such a rule). Rule 3.1' does not express any such natural process. It states only that a vowel of a certain, not phonetically defined type is fronted, unrounded,

and lowered after a palatalized consonant. This is the inevitable result of the use of ad hoc features like "alternating." With them anything is possible, nothing excluded, and no phonological rule is more or less natural than any other.

Let us consider an alternative way of describing the vowel alternation in Dialect B. We posit for this dialect the same five underlying vowels as for Dialect A, and we classify them according to the same specifications for Dialect A. Likewise we include in the grammar of Dialect B the distinctive features. Up to this point the grammars of the two dialects are identical, but we include additionally in the grammar of Dialect B *after* Rule 3.1 has operated the rule:

$$3.2 \quad \begin{bmatrix} -\text{consonantal} \\ +\text{vocalic} \\ +\text{back} \end{bmatrix} \rightarrow \begin{bmatrix} -\text{low} \\ +\text{round} \end{bmatrix} / \text{—} \begin{bmatrix} +\text{consonantal} \\ +\text{anterior} \\ -\text{coronal} \end{bmatrix}$$

(Any back vowel must be not low and must be round preceding labial consonants, i.e. a > o before labials.)

In Dialect B we will then have the derivations:

Underlying Forms:	ap	ap	op	op	ep	ep
Rule 3.1:	——	t,æp	——	——	——	t,ep
Rule 3.2:	op	——	op	t,op	——	——
Final Forms:	op	t,æp	op	t,op	ep	t,ep
Glosses:	'dog'	'to the dog'	'cat'	'to the cat'	'fly'	'to the fly'

The word /ep/ 'fly' with its dative form /t,ep/ has been included for the purpose of a later illustration. The grammar of Dialect A has only Rule 3.1; hence the output of this rule gives the proper forms of these words in Dialect A. The output of Rule 3.2, acting on the forms produced by the operation of Rule 3.1, gives the correct forms in Dialect B. For comparison we have:

Dialect A	*Dialect B*	
ap	op	'dog'
t,æp	t,æp	'to the dog'
op	op	'cat'
t,op	t,op	'to the cat'
ep	ep	'fly'
t,ep	t,ep	'to the fly'

What have we done in analyzing these two dialects and, more important, what have we not done? Our sole aim was to construct a grammar fragment that correctly accounts for certain kinds of vowel alternations in these two dialects. In Dialect A we found that five vowel phonemes are needed to specify the lexical shape of morphemes, and we have posited for that grammar a rule (3.1) that fronts /a/ after a palatalized consonant. We found that Dialect B also has in its inventory of systematic phonemes five underlying vowels—the same five as Dialect A with respect to their distinctive feature specifications—and we have included in the grammar of Dialect B a rule (3.2) that raises and rounds /a/ to [o] and that applies after Rule 3.1.

We have come closer to a treatment of dialect difference that in justice to the intuitive notion of dialect shows a great amount of relatedness between these two dialects: they share similar sets (in this case identical sets) of systematic phonemes; their grammars contain identical rules (Rule 3.1 and by assumption all the other phonological rules); and the final phonetic forms generated by the two grammars are similar and to some extent identical. This latter point is the most salient common-sense characterization of dialect relatedness: they sound pretty much alike, and there is considerable mutual intelligibility. Yet even though we have found that the grammars of the two dialects have much in common, there are differences between the two, which can be stated precisely. Using terms that dialectologists might traditionally apply, the isogloss between these two dialects is defined by a difference in their grammars: one dialect has a rule absent in the other.

There are several noteworthy things we did not do and indeed would not care to do. We did not assume that the two dialects were in some ways similar—that, for example, they had the same set of underlying vowels or that they shared a number of identical rules. Economy and the naturalness of rules (as well as considerations about what distinctive features should be present in a natural language) would induce us to posit an underlying /a/ in the grammar of Dialect B even though there is no [a] sound in the paradigm for 'dog' and irrespective of the fact that Dialect A has a systematic phonemic /a/ which happens to be realized as phonetic [a]. We were led to derive [æ] from underlying /a/ rather than /o/ or /ɔ/ because we could do so with fewer features and with natural rules. In principle, our linguistic theory would have led us to the same grammars even if we knew only of the existence of the one dialect and not the other. In practice, of course, our expectation of what to find in one dialect is conditioned by what we know of other related dialects, and we use anything we can get our hands on to come up with a good analysis. But in our role of linguist qua constructor of grammars we are concerned only with accounting for the intrinsic knowledge underlying speech in one dialect, and the existence of a second dialect is absolutely irrelevant to this goal. In short, the similarities (and divergences) between the two grammars are a result of purely synchronic analyses carried on for each dialect individually and are in no way an assumption upon which our analysis is based.

Finally, in regard to what was said earlier about overall patterns and diasystems, it is instructive to see what kind of dialect comparison emerges from application of one of these constructs. We have minimal pairs for /e/, /a/, and /o/ in Dialect A—[ep] : [ap] : [op]; and minimal pairs for /e/, /æ/, and /o/ in Dialect B—[t̪ep] : [t̪æp] : [t̪op]. Assuming in addition then that /i/, /u/, and /a/ are (autonomous) phonemic in both dialects, we might have the following inventories of autonomous (as distinct from systematic, or underlying) phonemes in the two dialects:

Dialect A   *Dialect B*

i		u	
e		o	
a			

i		u
e		o
æ		a

Following Weinreich we can establish the diasystem:

$$\big/\!\big/\ i \approx e \approx a \approx \dfrac{A/a \sim o/}{B/æ \sim o/} \approx o \approx u\ \big/\!\big/$$

This diagram says both too much and too little. It indicates much dissimilarity between the dialects in regard to the correspondences A/a ~ o/ : B/æ ~ o/, yet we have seen that the dialects correspond with total regularity in this area if we derive [æ] from an underlying /a/ in Dialect B and do not set it up as a separate phoneme. The diasystem does not inform us that there is a general rule predicting [æ] in both dialects. The diasystem does not inform us that the correspondences A/a ~ o/ : B/æ ~ o/ are not so much indications of structural (that is to say, phonemic) disparity between the dialects as they are a slightly peculiar set of partially phonetic, partially phonemic correspondences between them. Thus, while the diasystem does serve to elucidate some of the relationships between the dialects, it also renders them more disparate than the facts justify without telling us specifically where and how they differ.

Our example is hypothetical, but not atypical or pathological. A quite similar example from Russian dialects is cited in Halle (1962:69–70), and there seems to be no reason for assuming that comparable cases are unusual, quite the contrary. One example of this sort has been given by Lamb (1966:542) for Monachi, a Uto-Aztecan language of California. In the dialect of Bishop, California, there is an alternation between [m] and nasalized [w̃]: [w̃] occurs only after vowels, [m] never occurs after vowels, and both segments are always followed by vowels. Example: [miyawai] 'will go', [taw̃iyawai'na] 'our future going'. In the dialect of North Fork the corresponding alternation is between [m] and non-nasalized [w]: [miyawai], [tawiyawai'na], [tawiyawai'na] 'our future going'. There is in both dialects a phone [w] which does not alternate with [m] and is in contrast with it. Example: [wiya] 'acorns', [tawiya] 'our acorns'.

To account for this tiny portion of data in our grammar of the Bishop dialect we posit two systematic phonemes /m/ and /w/ classified by at least the features:

	w	m
vocalic:	−	−
consonantal:	−	+
nasal:	−	+
coronal:	−	−

And we include in the grammar of the Bishop dialect the rule:

3.3
$$\begin{bmatrix} + \text{nasal} \\ - \text{coronal} \end{bmatrix} \rightarrow [- \text{consonantal}] \, / \, V \text{\underline{\quad}}$$

(/m/ > [w̃] following a vowel.)
The North Fork dialect is similar in having underlying /m/ and /w/ as well as Rule 3.3, but its grammar has a rule that applies after Rule 3.3 and states that any glide (such as [w] or [w̃]) is not nasal:

3.4
$$\begin{bmatrix} - \text{vocalic} \\ - \text{consonantal} \end{bmatrix} \rightarrow [- \text{nasal}]$$

(A glide has the feature minus nasality.)
The difference between the two dialects in this one respect can be described by assuming the presence of a rule in the grammar of the North Fork dialect that is absent in the grammar of the Bishop dialect. The underlying segments for this little piece of the grammar are in both dialects the same, /w/ and /m/. (The analysis given here might require modification in the presence of additional data on the dialects. It might, for example, be economical to combine Rules 3.3 and 3.4 into a single rule. For further discussion see Section 3.3 under SIMPLIFICATION.)

The diasystem for this subpart of the autonomous phonemic systems of the two dialects would be // m ≈ w //, which reveals exactly nothing of the way in which the dialects differ. In autonomous phonemic terms the m : w̃ alternation in Bishop is allophonic since the phone [w̃] can be assigned to the phoneme /m/; in North Fork the m : w alternation is morphophonemic since a phone [w] must be assigned to the phoneme /w/ irrespective of whether it alternates under stable conditions with [m] or not. If we recognize an autonomous phonemic level intermediate between the systematic phonemic and phonetic levels in our grammar, then the two grammars differ at two levels: the morphophonemic and the allophonic. We have seen that the difference can be described simply as the addition of a single rule in a generative

grammar that posits no level of representation intervening between the systematic phonemic and the systematic phonetic. In particular, this example demonstrates that meaningful dialect comparison does not necessarily emerge from comparing the phonemic inventories of two dialects, for in this case the phonemic inventories—whether autonomous or systematic phonemic—are identical, and whatever differences there are arise via rules that state realizations of phonemes.

The moral to be drawn from this discussion is that to gain any insight into dialect differences we must concern ourselves with the *grammars* of languages, not their vowel or consonant systems, lists of morphemes, and so on. That is, the study of dialect differences is the study of how the grammars of the dialects differ. By implication, the study of linguistic change is the study of how the grammars of languages change in the course of time. We have nothing to gain from comparing phoneme inventories at two different stages of a given language and seeing what sound has changed into what other sound. Such a comparison gives as little insight into linguistic change as a comparison of before-and-after pictures of an earthquake site gives into the nature of earthquakes.

3.3 TYPES OF CHANGE

For the present we shall be concerned with drawing the consequences of the statement made earlier that all change can be traced to the situation in which two dialects of a language have become different. This includes the possibility that one of these dialects is the immediate chronological predecessor of the other, e.g. Old English and Early Middle English, or Late Proto-Indo-European and Early Proto-Germanic. To say that dialects have become different is to say that the grammars of these dialects are different, and we shall now discuss the ways in which this can happen. Our examples will be taken from phonology; syntactic change will be discussed in Chapter 6. The types of change we shall discuss in the remainder of this chapter are all what we will call *primary change* (change in the rule component) as distinct from *restructuring* (change in underlying representations), which will be discussed in Chapter 4.

RULE ADDITION. The examples of dialect differences given in Section 3.2 involve two dialects which differ by the presence in one of the grammars of a rule absent in the other. Transposed into terms of historical linguistics, this means that one of the ways for sound change to take place is by the addition of a rule to the grammar. Instances of this kind are commonplace, and many well-known sound changes are best described by rule additions.

Thus, the change of Vulgar Latin ū > Old French ü was presumably of this kind. To the grammar of Vulgar Latin was added the rule:

3.5
$$\begin{bmatrix} V \\ + \text{high} \\ + \text{long} \end{bmatrix} \rightarrow [-\text{back}]$$

(A vowel which is high and long becomes front; its roundness value is unchanged. Hence, high back round $\bar{u}$ becomes high front round $\bar{u}$.)

Palatal umlaut—the fronting of back vowels that are followed by i, $\bar{\imath}$, or j—in the Germanic languages is a second example of rule addition. There is no reason to assume that umlaut was present in the grammar of Gothic, the oldest attested of the Germanic languages. (Early Germanic loanwords in Finnish show no umlaut, e.g., Finnish *patja* 'mattress', not *pätjä* or *pejja*, from *badja*, cf. Gothic *badi* 'bed'.) We thus have Gothic *gasts : gasteis* 'guest (nom. sing. and plu.)', where *ei* spells [i:], and there is no documentary indication of umlaut in the form *gasteis*. In Old High German and Old Saxon, on the other hand, umlaut of short a is clearly present and indicated in the manuscripts, so that we have *gast : gesti* ($a > e$ under influence of the following i). Similarly, corresponding to Gothic *fotus : fotjus* 'foot (nom. sing. and plu.)' without umlaut in *fotjus*, we have Old English *fōt : fēt* and Old Norse *fōtr : fø̄tr*. Old Norse points up the older situation, in which $\bar{o}$ is umlauted to $\bar{ø}$ (spelled $\acute{o}$ in Old Norse) under the influence of a following j, which since had become lost; and most dialects of Old English have unrounded their front rounded (umlaut) vowels, thus $\bar{ø} > \bar{e}$. We account for this by assuming the addition of an umlaut rule to the grammar of a dialect (or dialects) of Germanic from which Old English, Old Saxon, Old High German, and Old Norse (but not Gothic) derive. This rule has the general shape:

3.6
$$V \rightarrow [-\text{back}] / \underline{\quad} C_1 \begin{bmatrix} -\text{consonantal} \\ +\text{high} \\ -\text{back} \end{bmatrix}$$

(A vowel is fronted when followed by one or more consonants plus a vowel or glide segment that is front and high, i.e. $\breve{u} \, \breve{o} \, \breve{a} > \breve{u} \, \breve{o} \, \breve{a}$ when followed by one or more consonants and $\bar{\imath}$ or j. Details of this rule would have to be modified in order to account for dialect-specific developments such as the fact that in the West Germanic dialects of Old High German and Old Saxon the umlaut of short a is e and not $ä$. Also, certain consonants and consonant clusters prevented umlaut from occurring in various dialects.)

The rule that subsequently unrounds front rounded vowels in most dialects of Old English can also be regarded as a case of rule addition. (It is better considered as simplification—the "activation" of the universal marking convention that the unmarked value of roundness for front vowels is [−round], cf. Chomsky and Halle 1968:405. For illustration, however, this change is treated here as rule addition.) In the early documents of Old English 'king'

is spelled *cyning* (y is high front round $\bar{u}$), and in the later documents *cining*, indicating that unrounding has taken place. Likewise, in the older Old English documents and regularly in Anglian manuscripts, isolated occurrences of $\bar{o}e$ are found for the palatal umlaut of $\bar{o}$, e.g. *bōc : bōec* 'book (nom. sing. and plu.),' but in the later writings we find only $\bar{e}$, as in *bōc : bēc* (cf. similar *foot : feet*). We assume that a rule requiring front vowels to be unrounded was added to the grammar of the unrounding dialects of Old English:

3.7
$$\begin{bmatrix} V \\ -\text{back} \end{bmatrix} \rightarrow [-\text{round}]$$

(Front vowels must be nonround, i.e. $\breve{u} \, \bar{o} > \breve{\imath} \, \breve{e}$. Note that Rule 3.7 applies *after* Rule 3.6.)

As a final example of rule addition let us take one step in the Germanic Consonant Shift (Grimm's Law), namely $b \, d \, g > p \, t \, k$. Corresponding to $b \, d \, g$ in most of the Indo-European languages, the Germanic languages have, in their earliest reconstructible stages at least, $p \, t \, k$. Examples: Greek *déka*, Latin *decem* versus Gothic *taíhun*, English *ten*; Greek *génos*, Latin *genus*, Sanskrit *janas* versus English *kin*, Old High German *kunni*, Old Norse *kyn*. We assume that a rule making stops voiceless was added to the grammar of the Northwesterly Indo-European dialect from which Germanic developed:

3.8
$$\begin{bmatrix} +\text{obstruent} \\ -\text{continuant} \end{bmatrix} \rightarrow [-\text{voice}]$$

(Any noncontinuant obstruent—a stop—must be voiceless.)

So far we have discussed rule additions with little reference to *where* the rule is added. In general we must be more precise than this, for it can make a difference at which point a rule is added to the grammar of a language, as is implicit in the notion that the phonological component of a grammar is constituted as a system of rules, some or all of which may be ordered with respect to each other. In all natural languages previously investigated it has been found to be the case that at least some phonological rules must be ordered with respect to each other.

As a concrete example, let us consider Rules 3.6 (umlaut) and 3.7 (unrounding) in Old English. Since Rule 3.6 (umlaut) is present in all of the early Germanic dialects except Gothic and sporadically elsewhere, we are quite certain that Rule 3.6 (umlaut) was added to the grammar of Old English earlier in time than Rule 3.7 (unrounding). The manuscript evidence supports this. Let us suppose then a stage of Old English the grammar of which contains Rule 3.6 (umlaut) but not Rule 3.7. We would then have such derivations as the following:

Base Form:	kuning	dūstig	hōrjan
Rule 3.6:	küning	dūstig	hōrjan
Early OE Spelling:	cyning	dystig	hōeran
Gloss:	'king'	'dusty'	'to hear'

(The -j- in hōrjan is lost subsequent to its triggering of umlaut.) If the unrounding rule (3.7) is added now, after Rule 3.6, we will obtain the correct forms for later Old English:

Base Form:	kuning	dūstig	hōrjan
Rule 3.6:	küning	dūstig	hōrjan
Rule 3.7:	kining	distig	hēran
Late OE Spelling:	cining	distig	hēran
Gloss:	'king'	'dusty'	'to hear'

If, conversely, Rule 3.7 were added to the grammar of Early Old English *before* Rule 3.6, then there would be absolutely no change in the final phonetic forms. Rule 3.7 states that front vowels must be nonround; thus Rule 3.7 will change segments only if the language has front rounded vowels. But since the only front rounded vowels in Old English arise through the operation of umlaut, Rule 3.7, if added prior to umlaut, would have no forms to change:

Base Form:	kuning	dūstig	hōrjan
Rule 3.7:	....	dūstig	hōrjan
Rule 3.6:	küning	dūstig	hōrjan
Phonetic Form:	küning	dūstig	hōrjan
OE Spelling:	cyning	dystig	hōeran

If, therefore, Rule 3.7 had been added to the grammar of Early Old English before Rule 3.6, we could not obtain the correct Late Old English forms *cining*, *distig*, and *hēran*.

In most cases of rule addition where we have all the necessary documentary and comparative information to determine rule chronologies with precision, the rules are added late in the grammar, at the end of the phonological component. This is not necessarily universal, however; it is entirely possible that our impression that rules tend to be added relatively late rests on insufficient evidence: few languages have written records, and even in those few that do we seldom can determine beyond question the relative chronology of two rules. We are lucky in the case of Old English, which has extensive documentation from circa A.D. 700, and in a few other cases from Indo-European languages.

One of the more certain cases of the addition of a rule not at the end of the phonological component is provided by Lachmann's Law in Latin (extensively discussed in Kiparsky 1965). Indo-European had a regressive voicing assimilation rule of the form:

$$3.9 \quad [+\text{obstruent}] \rightarrow [\alpha \text{ voice}] / \underline{\hspace{1em}} \left[\begin{array}{c} +\text{obstruent} \\ \alpha \text{ voice} \end{array} \right]$$

(An obstruent takes on the voicing value of the immediately following obstruent: it is voiceless if the following obstruent is voiceless, it is voiced if the following obstruent is voiced.) Thus, via Rule 3.9, to the Sanskrit root *vid-* 'to know' we have *vēt-tha* 'you know' from underlying /vēd-tha/; to the Greek root *leg-* (*légō* 'I say') we have *lek-tó-s* 'gathered' from underlying /leg-tó-s/; and to the Latin root *scrīb-* 'to write' we have *scrip-s-ī* 'I wrote' and *scrip-t-um* 'having been written' from underlying /skrib-s-ī/ and /skrib-t-um/. In these cases voicelessness is assimilated regressively, but voicing too could be assimilated (α is + in Rule 3.9): to the Sanskrit root *śak-* 'to be able' we have the second plural middle imperative form *śagdhvam* from underlying /śak-dhvam/. Rule 3.9 was inherited in the grammar of Latin, where it is needed as a synchronic rule to account not only for forms such as *scrībō : scrīpsī : scrīptum* but also for those like *appellō* 'I call' from underlying /ad-pellō/ and *accipiō* 'I receive' from underlying /ad-kapiō/.

Latin has quantity alternations among vowels of the following sort: *agō* 'I drive, lead'; *āctum* 'having been driven, led'; and *regō* 'I rule': *rēctum* 'having been ruled'. Yet this short : long vowel alternation is not found in other forms such as *faciō* 'I make': *factum* 'having been made'; and *capiō* 'I take': *captum* 'having been taken'. The solution seems clear. The underlying forms of *āctum*, *rēctum*, and so on are /agtum/, /regtum/ (cf. *agō, regō*), and there is a rule that lengthens vowels when followed by an obstruent cluster, the first member of which is voiced, though analogy has disturbed the effects of this rule. The underlying forms of *factum, captum,* and so on have only voiceless obstruents (i.e. /faktum/, /kaptum/) since no other forms in the paradigm have voiced obstruents, and this rule will not lengthen vowels in these words. The rule, *Lachmann's Law*, can be stated as follows:

$$3.10 \quad V \rightarrow [+\text{long}] / \underline{\hspace{1em}} \left[\begin{array}{c} +\text{obstruent} \\ +\text{voice} \end{array} \right] \left[\begin{array}{c} +\text{obstruent} \\ -\text{voice} \end{array} \right]$$

(A vowel is long before the sequence voiced-obstruent-plus-voiceless-obstruent.)

Now the question is: "Where was Rule 3.10 added in the grammar of Latin?" We might first assume that it was added at the end of the

phonological rules. In this case it would apply *after* Rule 3.9, the voicing assimilation rule, since the latter rule is an old rule inherited from Indo-European and since Rule 3.10, Lachmann's Law, is a Latin innovation not shared by the other Indo-European languages. This assumption would give derivations such as the following:

Base Form:	agō	agtum	fakiō	faktum
Rule 3.9:	———	aktum	———	———
Rule 3.10:	———	aktum	———	———
Final Form:	agō	aktum	fakiō	faktum
Latin Spelling:	agō	actum	faciō	factum

We see that this assumed order of the two rules does not give the right results for forms such as *āctum, rēctum, lēctum* 'having picked out' (cf. *legō*). The point is that if Rule 3.10 is added at the end of the grammar, then it must apply vacuously (i.e. without effect) since Rule 3.9 will already have made obstruent clusters like the -*gt*- in *agtum* voiceless throughout. In this event, the structural analysis of Rule 3.10 will never be met, and the rule will never apply. If, on the other hand, we assume that Lachmann's Law was added as a rule of grammar *prior to* Rule 3.9 (voicing assimilation), then we will obtain the following correct results:

Base Form:	agō	agtum	fakiō	faktum
Rule 3.10:	———	āgtum	———	———
Rule 3.9:	———	āktum	———	———
Final Form:	agō	āktum	fakiō	faktum
Latin Spelling:	agō	āctum	faciō	factum

A second example of nonchronological rule addition may be taken from the history of German, where the sequence [xs] became [ks] within morphemes: *oxs > oks* 'ox', *zexs > zeks* 'six', *laxs > laks* 'salmon'. However, [x] remained when separated from [s] by a morpheme boundary: *maxst* 'you make', *ne:xst* 'next', *laxst* 'you laugh'. Accounting for this by rule addition at the end of the phonological component would require a rule of the sort:

xs > ks only if no morpheme boundary intervenes

This rule violates the apparently valid empirical hypothesis (Chomsky and Halle 1968:364) that processes operating within morphemes normally also apply across morpheme boundaries. The simpler explanation is that this change

was added as a morpheme-structure constraint—not, in other words, at the end of the phonological component, but as a condition on the configuration of morphemes.

As these examples show, one must not facilely make the assumption that rules are added only at the end of the phonological component. Perhaps that is the more common occurrence (though this is only an assumption), and for reasons that have to do with the disruption of communication. As Halle (1962:66) has pointed out, "Language change is normally subject to the constraint that it must not result in the destruction of mutual intelligibility between the innovators—i.e. the carriers of the change—and the rest of the speech community." All things being equal, a rule that is added late in the grammar will disrupt communication less than a rule added early in the grammar, though in some cases it will make no difference. Such was the case in Old English discussed earlier: if Rule 3.7 (unrounding) had been applied early—specifically before Rule 3.6, the umlaut rule—then it would not change the previous output of the grammar in the slightest.

It can, however, make a very great difference, depending on the grammar and the rule which is inserted. Suppose, for example, that a rule raising [æ:] > [e:] were added at the end of the phonological component of American English. Instead of [bæ:d] 'bad' and [kæ:n] 'can', we would say [be:d] and [ke:n]. Communication would hardly be impaired at all. In fact, there are many speakers of American English, located primarily in New York City, who have done just this (Labov 1965:102-113), and a great many immigrants pronounce [æ] as [e:], yet those of us who don't do this have no trouble talking to those who do. It is simply registered as some kind of accent: urban New York, German immigrant, or whatever.

If this rule had been inserted very early in the set of phonological rules, the effect it would have on the final phonetic output of the grammar would be much greater. There is in the grammar of English a rule raising underlying /æ/ to [ē] (the macron denotes tenseness, or length). Thus, the underlying forms of *compare* and *comparison* [kʰʌmpe:r] and *comparison* [kʰʌmpærəsin] contain [æ], and the vowel-raising rule operates on the /æ/ in *compare* to give [e:]. Also, in its present-day grammar English has a vowel shift rule that (among other things) raises underlying /ē/ to [i]. This rule is needed to account for such alternations as *keep* [kʰi:p] : *kept* [kʰɛpt], both of which have underlying /ē/. The laxing rule gives /kēp/ > [kept] [kʰɛpt] which later rules take to [kʰɛpt], and the vowel shift rule gives /kēp/ > [kip] (> [kʰi:p]). (Note that here and elsewhere the vowels in English *beet, bait, pod, boat, boot*, and so on are written as simple vowel plus [:]. This implies not that "length" is the only relevant feature phonetically present but that a variety of phonetic features such as tenseness, length, and offglide are involved.) For the details of this analysis see Chomsky and Halle (1968).

Now let us suppose that the rule raising /æ/ to [ē] were added at the very beginning of the phonological component. We would then have

Underlying: ē   ǣ
                       ......
         ǣ > ē   ē
Vowel Shift: ī
                ī

Under these conditions *compare* would come out [kʰampiːr] and *comparison* would come out [kʰampɛrəsin]. The degree to which changes of this sort would adversely affect communication is open to speculation and could be tested experimentally, but clearly adding the rule /ǣ/ > [ē] at the beginning of the phonological component has brought about a greater deviation from normal pronunciation than results when the same rule is added at the end of the phonological rules.

At present not a great deal is known about the "disruption of mutual intelligibility" criterion. We know that some such tolerance point exists; otherwise we would expect to find cases of radical communication breakdown between speakers belonging to successive generations. But just how to formulate a formal constraint that captures the notion of a point at which mutual intelligibility is disrupted by change is neither easy nor obvious. (It may well be that this constraint should not be stated as a constraint on the grammar at all, but rather should be accounted for elsewhere in the theory.) Languages, or rather their speakers, seem to be able to tolerate seemingly radical changes without slackening their stride to any great extent, yet we know of no language that anywhere in its history has undergone really pathological changes such as "All high vowels become mid," "All high vowels become low, all front vowels become back, and all back vowels become mid."

In any case, though evidence is not conclusive, it is plausible to assume that rules tend to be added at the end of the phonological component rather than earlier because communication is thereby less affected. Yet our impression that late rule addition is statistically favored may be due not to some universal principle but merely to insufficient knowledge of sound changes. Numerous instances of rule insertion at points other than the end are attested. Others are shown for Mohawk and Oneida by Postal (1968:245-260).

Before leaving rule addition, it should be observed that this kind of primary change corresponds to what has traditionally been known as *innovation*. Each case presented—Vulgar Latin *ū > ü*, Germanic umlaut, Grimm's Law *b d g > p t k*, and Lachmann's Law—falls in the category of innovations in the individual languages.

RULE LOSS. Another kind of primary change can be deduced from the fact that grammars of dialects sometimes differ by the presence or absence of a single rule: it may be that a rule has been lost from the grammar. We shall discuss two such cases here—one from Yiddish, the other from Gothic. We begin with the Yiddish example since the spoken language is still available to us. (Cf. Kiparsky 1965 and Kiparsky 1968b.)

Middle High German, from which Yiddish dialects derive ultimately though not directly, had a rule that devoiced final obstruents; we posit this rule on the basis of Middle High German alternations such as *gap* 'he cut'. In word-final position the contrast between voiced and voiceless stops (and probably fricatives too, though the orthography is less clear on this point) is neutralized in favor of the voiceless member. The obvious way to handle such alternations is to posit underlying voiced obstruents in the forms involved and include a terminal devoicing rule in the grammar. The underlying representations of the forms just cited would then be /gab : gāben, tag : tage, sneid : snīden/, and the following rule would convert word-final voiced obstruents to voiceless ones:

3.11 [+ obstruent] → [− voice] / ___ #

(Obstruents are devoiced word-finally.)

Earliest attested Old High German, the predecessor language of Middle High German, had no such rule in its grammar; the Old High German forms of the words cited were *gab : gābum(es)*, *tag : taga*, *sneid : snīdan*. It is clear from the written records that Rule 3.11 was added to the grammar of Old High German between A.D. 900 and 1200 depending on the dialect. The terminal devoicing rule is present in the vast majority of the modern German dialects, including Standard German, though in some dialects it is limited to word-final fricatives. Some German dialects, however, do not have such a rule in their synchronic grammars, and in particular many Yiddish dialects do not have this rule. We may cite examples from Standard Yiddish, whose grammar lacks Rule 3.11 in any form: *hob* 'I have': *hobm* 'we have', *lid* 'song': *lider* 'songs', *tog* 'day': *teg* 'days', *noz* 'nose': *nezer* 'noses', *rov* 'rabbi'.

Two explanations of this are possible. One, which we reject, is that Yiddish never added to its grammar a rule that devoiced final obstruents. Evidence still in the language argues against this, for Standard Yiddish has numerous words which show that Rule 3.11 was operative in the language at some earlier stage: *avek* 'away', *hant* 'hand', *gelt* 'money'. All of these words had underlying *voiced* obstruents earlier, cf. the Middle High German cognates *hant : hende, wec : weges, gelt : geldes*. The final *voiceless* obstruents in the Yiddish forms could only have resulted from a stage in the development of Yiddish when Rule 3.11 was present in the grammar. In addition to purely internal evidence such as this, there is direct textual evidence for such a rule in a Yiddish rhyme of the thirteenth century (Röll 1966). This rhyme begins with the phrase *gūt tak* in Hebrew letters where the *k* in *tak* is spelled with the Hebrew letter for *k* (kuf) and not *g* (gimel). The Standard Yiddish expression for this is *a gutn tog*, cf. Standard German *Guten Tag* 'good day'. Furthermore, many Yiddish dialects still maintain final devoicing (Herzog

1965:220-223). This evidence, taken together, leads us to reject the proposal that Yiddish never had a terminal devoicing rule in its grammar.

One might, of course, try to account for the presence versus the absence of devoicing in Yiddish dialects by appeal to borrowing or to areal influence. There are insurmountable difficulties in such explanations, as Weinreich (1963) has demonstrated, and the various alternative explanations will not be investigated here.

It is more reasonable to assume that the earliest Yiddish dialects had in their grammars Rule 3.11 as an inheritance from Middle High German, but that most of the dialects since lost this rule from their grammars. The underlying voiced final obstruents in *tog, hob, lid, noz,* and so on have been carried along unchanged through the lexicons of successive generations of Yiddish speakers. As long as Rule 3.11 was present in the grammar to act upon these forms, they would have voiceless final obstruents in their phonetic realizations —*tok, hop, lit, nos,* and so on—rather like Standard German *Tag* [tʰaːk], *hab'* [haːp], *Lied* [liːt]. With the loss of Rule 3.11 the underlying forms come through unaltered as regards their final obstruent; that is, voiced word-final obstruents at the underlying level are realized phonetically as voiced.

Instances of rule loss from a grammar are by no means uncommon. As we shall see in Chapter 4, rule loss is concomitant with the type of change we call restructuring, and restructuring is frequent enough in the history of languages. Let us investigate another case of simple rule loss where the evidence is reasonably clear.

All the early Germanic dialects except Gothic have an original alternation between voiceless and voiced fricatives that shows up with particular regularity in the principal parts of strong verbs. This phenomenon is known as *grammatical change.* It is a result of Verner's Law, which states that "Germanic voiceless spirants remained voiceless if the preceding syllable had the Indo-European accent, but became voiced in voiced surroundings if the preceding syllable had been unstressed in Indo-European times" (Prokosch 1939:61). Verner's Law may be stated as:

$$3.12 \quad \begin{bmatrix} + \text{obstruent} \\ + \text{continuant} \end{bmatrix} \rightarrow [+ \text{voice}] \, / \, \begin{bmatrix} + \text{voice} \\ - \text{accented} \end{bmatrix} \underline{\quad} [+ \text{voice}]$$

(Fricatives become voiced in voiced surroundings following an unaccented segment.) Examples of grammatical change are:

	Inf.	Past Sg.	Past Pl.	Past Part.	Gloss
OE:	snīþan	snāþ	snidon	sniden	'to cut'
ON:	kiósa	kaus	kørom	kørenn	'to choose'
OS:	tiohan	tōh	tugun	gitogan	'to pull'
OHG:	ziohan	zōh	zugum	gizogan	'to pull'

In the early dialects of Germanic, stress is localized on the root syllable, so that the cited forms reflect the place of original Indo-European accent only in alternations of the sort þ : ð, s : r, h : g, which show secondary changes from the original Proto-Germanic set of alternations which were voiceless : voiced, e.g. þ : ð, s : z, x : g. To see the reason for such differential treatment in the principal parts of verbs, consider these same principal parts as they would have appeared in Indo-European, e.g. IE *dewk-* 'to pull' (Latin *dúcō*):

	Inf.	Past Sg.	Past Pl.	Past Part.
IE:	déwkono	dówke	dwkṇt	dwkóno
Early Proto-Gmc.:	téwxana	táwxe	tuxínþ	tuxána
Late Proto-Gmc.:	téwxan	táwx	túgun	túgun

Here, the difference between "Early Proto-Germanic" and "Late Proto-Germanic" is that the former stage still has Indo-European accent placement, while the forms in the latter stage have undergone Verner's Law (Rule 3.12) and the accent (or stress) has shifted uniformly to the root vowel.

In order to account for these alternations in the synchronic grammars of the Germanic dialects (not including Gothic), we need a rule that voices fricatives in the requisite environments. It is no longer possible to assume that Rule 3.12 is present in these synchronic grammars, for there is no motivation for assuming different placements of the accent in the principal parts of verbs, and Rule 3.12 crucially requires us to accent the infinitive (and all present tense forms) and the past singular forms on the root vowel, and to accent the past plural and the past participle on the suffix vowel. For these reasons we include in the grammars of the earliest Germanic dialects the rule:

$$3.13 \quad \begin{bmatrix} + \text{obstruent} \\ + \text{continuant} \end{bmatrix} \rightarrow [+ \text{voice}] \, / \, \begin{bmatrix} + \text{Past Plural} \\ + \text{Past Participle} \end{bmatrix} \, +]$$

(Fricatives are voiced in the past plural and past participle when stem-final.)

We may regard Rule 3.13 as the altered synchronic survival of the Germanic innovation (i.e. rule addition) of Rule 3.12, which is Verner's Law. Thus, starting with /sniþ-/ 'to cut' in the lexicon of Old English, the voiceless fricative þ would remain unchanged in the present and past singular, but would be voiced in the preterite plural and preterite participle—giving the correct forms *sniþan snāþ snidon sniden* (a rule applying subsequent to Rule 3.13 would convert the ð from þ into the stop *d*).

We have been careful to exclude Gothic from all of the foregoing comments, for Gothic alone among the Germanic dialects does not have grammatical change in the principal parts of its strong verbs. For the three verbs 'to cut', 'to choose', and 'to pull', cited earlier in their Old English, Old Norse, Old Saxon, and Old High German shapes, Gothic has:

Inf.	Past Sg.	Past Pl.	Past Part.
sneiþan	snaiþ	sniþum	sniþans
kiusan	kaus	kusum	kusans
tiuhan	tauh	tauhum	tauhans

with voiceless fricatives throughout.

What are we to assume has happened? Internal evidence from Gothic inclines us strongly to the assumption that the grammar of Gothic once included Rule 3.12 (Verner's Law) and likely for a time its synchronic guise Rule 3.13, but that the rule was deleted from the grammar. There are relic forms in the language that point to an original voiceless : voiced alternation precisely of the kind found elsewhere in Germanic: examples are *aih* 'I possess', *weihan* 'to fight': *du wigana* 'to the battle', *frawairþan* 'to perish': *frawardjan* 'to ruin', *wisan* 'to feast': *wizon* 'to revel', *þarf* 'I need': *þaurbum* 'we need', *filhan* 'to hide': *fulgins* 'hidden'. In these data there are enough past plurals (*aigum*, *þaurbum*) to justify the assumption that Rule 3.13 was once in the grammar of Gothic, presumably as part of its common Germanic heritage, but that it was lost. We assume that while this rule was part of the grammar, past plural and participle forms containing underlying voiceless fricatives were realized with voiced fricatives: e.g. *sniþum sniðans, kuzum kuzans, taugum tauʒans* parallel to *aigum* and *þaurbum*. But this was at the level of surface phonetic realization. The underlying systematic phonemic forms retained voiceless fricatives. With the loss of Rule 3.13, the great majority of strong verbs came to show at the phonemic level the structure all along present at the systematic phonemic level—voiceless fricatives in all the principal parts.

The relic forms are those in which the morphophonemic identification of related forms had presumably become so weak that a minor sort of restructuring had occurred in the lexicon. For example, *du wigana* 'to the battle', though originally derived from the verb *weihan* 'to fight', was no longer associated synchronically with it in the same close morphophonemic relationship as principal parts of strong verbs. Hence the *wigana* of *du wigana* was changed in the Gothic lexicon from original /wihana/ to /wigana/; similarly for the other relic forms, such as *wizon* 'to revel', which had lost its original transparent relationship with its lexical source *wisan* 'to feast' and had been altered in the lexicon to /wizon/ with /z/ in place of /s/. Two of the verb forms still demonstrating the original phonetic alternations—*aih : aigum, þarf : þaurbum*—have defective distributions in Gothic (and in the other Germanic dialects) as well as an aberrant set of endings. Both these verbs belong to the fossil class of "preterite-present verbs" from which we get our modal verbs in English: *can, may, dare,* and so on. In the third verb in Gothic with a remnant of the original distribution of voiceless : voiced fricatives—*filhan* 'to hide': *fulgins* 'hidden'—the form *fulgins* had come to be regarded as an

adjective rather than the past participle of *filhan*, and had been restructured in the lexicon to a separate adjective entry with /-g-/ no longer derived from the verb *filhan* with phonemic /-h-/. Similarly for *frawairþan : frawardjan*.

On balance, the evidence of the relic forms in Gothic points strongly to a rule loss. Such relic forms are our best evidence in making a case for loss of a rule, just as the relic form *avek* 'away' in Yiddish supports the assumption that the terminal devoicing rule was lost in that language. In this case, the adverb *avek* had been dissociated from its historical source *veg* (with retained morphophonemic final g) and restructured in the lexicon to /avek/.

To be sure, the claim for restructuring rests on reasonable probability, not certain fact: no one knows for sure what took place in a Yiddish or Gothic speaker's lexicon. But one is usually safe in appealing to restructuring when the process of deriving one form from another cannot be synchronically motivated as a rule for the grammar in question, yet the two forms are known to be related etymologically. Gothic *filhan* and *fulgins* are known to be from the same source in pre-Gothic. Yet one cannot motivate a rule for the synchronic grammar of Gothic which would derive adjectives from verbs, among them *fulgins* from *filhan*. Presumably the speaker of pre-Gothic derived *fulgins* and *filhan* from a single lexical source, as English speakers do for *divine* and *divinity*; but the speaker of recorded Gothic learned two separate lexical entries, much as we learn *drink* and *drench* as separate lexical items, even though the two have the same etymological origin.

RULE REORDERING. Another way dialects differ is in the ordering of certain of their rules. Thus, Dialect A contains in its grammar rules X and Y, which must apply in the order X first and Y second. Dialect B contains the same two rules but in the opposite order: Y first and X second. If the rules are crucially ordered in both dialects, a difference of output results. The number of attested examples of such reorderings increases as more and more languages are investigated from a generative point of view; and our present knowledge of relatively few reorderings indicates not so much their infrequency as the facts (1) that we have detailed histories of relatively few languages and (2) that researchers have not in general been on the lookout for reorderings.

Nevertheless, cases where two or more dialects differ in the order of application of the same or similar rules have been found for American English by Keyser (1963:310–311), for modern Rumanian by Vasiliu (1966), for Swiss German and Finnish by Kiparsky (1965 and 1968b), and for modern German dialects by Becker (1967:87–92). Corresponding to the synchronic cases, a number of instances of rule reorderings in historical linguistics have recently been unearthed, and we shall analyze two of these here.

The phonology of Modern Standard German contains two rules of interest here: one is a terminal devoicing rule (given already as Rule 3.11); the other lengthens vowels followed by voiced obstruents. The latter rule expresses a generalization about German phonology with only a few exceptions, such as

Ebbe 'low tide', *Widder* 'ram', and *Egge* 'harrow', mostly of Low German origin. This rule can be stated as:

3.14 $V \rightarrow [+\text{long}] \, / \, \underline{\quad\quad} \begin{bmatrix} +\text{obstruent} \\ +\text{voice} \end{bmatrix}$

(A vowel must be long before a voiced obstruent.)

Rule 3.14 accounts for the long vowels in such words as *Hagel* 'hail', *sagen* 'to say', *Tage* 'days', *fragt* 'he asks'. No prediction of vowel length can be made before voiceless obstruents, where both long and short vowels occur freely: *Betten* [bɛtən] 'beds', in contrast with *Beeten* [beːtən] 'beetroots'.

Rule 3.11, which devoices obstruents in word-final position, accounts in present-day German for voiceless : voiced phonological alternations of the following sort: [loːp]: [loːbəs] = *Lob* : *Lobes* 'praise, of praise'; [raːt] : [raːdəs] = *Rad* : *Rades* 'wheel, of the wheel'; [graːs] : [graːzəs] = *Gras* : *Grases* 'grass, of the grass'. [veːk] : [veːgə] = *Weg* : *Wege* 'path', 'paths'; [graːs] : [graːzəs] = *Gras* : *Grases* 'grass, of the grass'. For such morphemes with allomorphs differing in voice in the final obstruent, we set up base forms containing final voiced obstruents, and those in word-final position will correctly be devoiced by Rule 3.11. Thus, from /lob/ we would have among the formatives that are input to the phonological rules /lob/ 'praise' and /lobas/ 'of praise'; the former would become [loːp] by Rule 3.11, the latter would remain [loːbas] since /b/ here is not in word-final position. (Cf. the discussion in Section 2.2.)

The lexical entry for 'praise' will then be /lob/, similarly /rad/ 'wheel', /veg/ 'path', /graz/ 'grass', where the vowels are unspecified for length. This solution requires the two rules 3.11 and 3.14 to be crucially ordered: Rule 3.14, for vowel lengthening before voiced obstruents, must apply *before* Rule 3.11, for terminal devoicing. We will have the typical derivations:

Underlying Forms:	lob	lobas	veg	vega	graz	grazas
Vowel Lengthening:	lo:b	lo:bas	ve:g	ve:ga	gra:z	gra:zas
Final Devoicing:	lo:p	......	ve:k	......	gra:s	......
Phonetic Shape:	lo:p	lo:bas	ve:k	ve:ga	gra:s	gra:zas

(The underlying forms cited in these derivations are intended to facilitate presentation of the analysis at hand. They would not be the correct base forms if we were attempting to account for more of the generalizations in German phonology. In particular, /ə/ is not among the systematic phonemes of German —it is a reduction form of other vowels. Likewise, [v] is derived from underlying /u/, so that *Weg* 'path' has the lexical representation /ueg/.)

The order of these two rules in the synchronic grammar of modern German is not their chronological order. A grammar at an earlier stage in the history of German had the two rules crucially ordered in precisely the opposite

order. We know this because the reordering occurred within the period of written records, because certain archaic dialects have preserved the original ordering, and because relic forms in the standard language reflect the earlier order.

As was mentioned earlier, final devoicing (Rule 3.11) was an innovation in the grammar of most German dialects around A.D. 1000, in any case not later than 1200. Lengthening of vowels before voiced obstruents was an innovation in the grammar of Early Modern German; that is, the documents indicate that it was a rule added around A.D. 1400, several centuries later than the final devoicing rule was added. Grammars of German immediately subsequent to this, in the fifteenth and sixteenth centuries and even later, had the two rules in the chronological order of their addition: Final Devoicing followed by Vowel Lengthening. In such a grammar the derivations given earlier would have different phonetic outcomes:

Underlying Forms:	lob	lobes	veg	vege	graz	grazas
Final Devoicing:	lop	......	vek	......	gras	......
Vowel Lengthening:	......	lo:bas	......	ve:ga	......	gra:zas
Phonetic Shape:	lop	lo:bas	vek	ve:ga	gras	gra:zas

Comparing these final phonetic shapes with those given earlier for the grammar of modern German, we see that certain changes have taken place: some short vowels have become long, e.g. *lop > lo:p, vek > ve:k, gras > gra:s.* In traditional presentations this change would be called *analogical leveling*, here leveling under pressure from other forms in the paradigm that have long vowels. In Chapter 5 we shall discuss the problem of analogy in detail, and in Chapter 4 we shall ponder some of the factors that may have motivated the reversal of rule order. For the present we observe simply that we have here in the history of a single language two stages with rules identical but ordered oppositely.

We might entertain an alternative explanation to rule reordering, one which would fall in the category of RULE ADDITION. Is it not possible that Rule 3.11 (Final Devoicing), similar to the insertion of Lachmann's Law into the grammar of Latin at a point not reflective of its chronological position?

In general it is not always possible to determine whether one is dealing with a case of reordering or of rule insertion at an earlier point in the grammar. In this case we can settle the dispute in favor of reordering. In the first place, there are dialects still spoken in this century that preserve the original order of the rules (Behaghel 1928:276). There are now only a few, it is true, and they are receding, but their indisputable presence is living testimony to earlier grammars in which Terminal Devoicing synchronically preceded Vowel Lengthening.

Second, if Vowel Lengthening had been inserted in the grammar ahead of Terminal Devoicing, then derivations like those that produced:

lop   lo:bas   vek   ve:ga   gras   gra:zas

could never have existed. We would have had from the moment Vowel Lengthening was inserted only the derivations producing the modern forms:

lo:p   lo:bas   ve:k   ve:ga   gra:s   gra:zas

However, relic forms from the earlier derivations exist. One is the adverb *weg* [vek] 'away!' Presumably, this form was dissociated from its original noun source *Weg* [ve:k] 'path' at a time when [vek] was the only form possible for its source, i.e. at a time when the paradigm had the form *vek ve:ga* instead of the modern *ve:k ve:ga*. Another relic form is *ap* 'from' with short [a] from Middle High German *abe ~ ap* 'off'.

We conclude then that this example is a case of rule reordering and not of rule insertion at a point other than the end of the grammar. We cannot always enjoy such certainty, as the following example illustrates. The grammars of all the early Germanic dialects contained Rules 3.15 and 3.16 in approximately the shapes shown:

3.15
$$\begin{bmatrix} + \text{obstruent} \\ + \text{voice} \\ - \text{strident} \\ \alpha \text{ anterior} \\ \beta \text{ coronal} \end{bmatrix} \rightarrow [- \text{continuant}] \Big/ \left\{ \begin{matrix} \# \\ [+\text{nasal}] \end{matrix} \right\} \underline{\quad} \begin{bmatrix} + \text{obstruent} \\ + \text{voice} \\ - \text{strident} \\ \alpha \text{ anterior} \\ \beta \text{ coronal} \end{bmatrix}$$

(The voiced, nonstrident fricatives *b ð g > b d g* word-initially, after nasals, and in gemination. *Stridency* distinguishes *s* from *b ð g*. [+nasal] is the natural class of nasal consonants /m n/. In stating the third sub-environment of Rule 3.15 a convention proposed by Bach 1968 is used. This convention proposes that the two rules A → B / C ___ and A → B / ___ C be collapsed into a single rule written A → B / C with the environment bar deleted. Such a rule effectively says "A becomes B either before or after C." The sequences *bb ðð gg* become by two successive applications of Rule 3.15 first *bb ðð gg*, then *bb dd gg*.)

3.16
$$\begin{bmatrix} + \text{obstruent} \\ + \text{continuant} \end{bmatrix} \rightarrow [- \text{voice}] / \underline{\quad} \#$$

(Fricatives are devoiced word-finally.)

Rule 3.16 is a less general version of the terminal devoicing rule discussed already as Rule 3.11. It is needed in the synchronic grammars of the Germanic dialects to account for such alternations as Gothic *beidan baiþ bidum bidans* 'to await (principal parts)', where *d* is presumably a fricative, and Old Saxon *hof : hobos* 'court, courts'. Rule 3.16 also provides for alternations between fricatives and stops of the sort illustrated by OE *hæbbe* OS *habbiu* 'I have': OE *hæfde* OS *habda*, where Old English *f* before *d* was [b]. (Additional motivation for these rules is found in King 1968, where this case of reordering is analyzed in detail.)

The grammars of Gothic, Old Norse, and Old High German contain Rules 3.15 and 3.16 in that order. A derivation of four forms of the Gothic verb 'to bind' in the present indicative, second and third person singular, second and third person plural follows:

Underlying Forms:	binðiz	bindiz	binðeð	bindanð
Rule 3.15:	bindiz	bindiz	bindeð	bindand
Rule 3.16:	bindis	bindiþ	bindeþ	————
Phonetic Forms:	bindis	bindiþ	bindeþ	bindand

(A later rule would produce correct 2. pl. *bindiþ* from above *bindeþ*.)

In the group of Germanic languages comprised by Old English, Old Saxon, and Old Frisian, however, the opposite order of the two rules is required: Rule 3.16 must apply before Rule 3.15. In these languages we would have the derivations

Underlying Forms:	binðiz	binðiz	binðeð	bindanð
Rule 3.16:	binðis	binðiþ	binðeþ	bindanþ
Rule 3.15:	bindis	bindiþ	bindeþ	bindanþ
Phonetic Forms:	bindis	bindiþ	bindeþ	bindanþ

The correct final forms for the individual languages would be produced by rules specific to those grammars. In all of these three languages 3. pl. *bindanþ* was regularly lost before voiceless fricatives: 3. pl. *bindanþ > bindaþ*. This form was then extended over the entire plural (syncretism). In Old English unstressed *i* was early reduced to *e*, giving Old English *bindes bindeþ bindeþ bindaþ*.

This example illustrates that the grammars of one group of Germanic languages require these two rules in one order, the grammars of the other group require them in the opposite order. We cannot be sure whether there has been a reordering or whether Rule 3.16 was placed into the grammar of Old English, Old Saxon, and Old Frisian ahead of Rule 3.15. We are relatively

certain that Rule 3.15 preceded Rule 3.16 in time as an innovation in Germanic. In view of the fact that the Germanic languages once (perhaps around the birth of Christ) were a close-knit, homogeneous speech community located in southern Scandinavia, we would incline to the view that the grammar of the proto-language once contained the two rules in the chronological order 3.15 followed by 3.16, and that while the grammars of Gothic, Old High German, and Old Norse continue this original order, reordering took place in the grammar of the language from which developed Old English, Old Saxon, and Old Frisian. We shall see in Chapter 4 that there are perhaps independent grounds for assuming that reordering has taken place in this instance, but for the present we only observe that reordering seems likely though we lack definitive evidence. The desired evidence here would be written records before and after the reordering, relic forms, and observable dialects historically descended from Old English, Old Saxon, or Old Frisian that still maintain the original order.

We have postponed until Chapter 4 discussion of possible reasons why reorderings take place. It may be pointed out here, however, that different orderings of rules among neighboring dialects may in some instances be due to a kind of wave effect. That is, rules spread out from prestige core dialects, and different rates of diffusion of the rules lead to different orderings.

Consider hypothetical dialects A and B. Dialect A has a Rule 1 and not a Rule 2 in its grammar, whereas Dialect B has Rule 2 and not Rule 1 in its grammar, and the contiguous dialects C and D have neither rule. This situation may be schematically represented by the following figure:

Now assume that Rules 1 and 2 diffuse away from their original areas towards Dialects C and D, that speakers of Dialects C and D borrow Rules 1 and 2 into their grammars, adding them on at the end. Further assume that Rule 1 is borrowed immediately from Dialect A into Dialect D but later into Dialect C, and that Rule 2 is transmitted rapidly from Dialect B into Dialect C but later into Dialect D. This could give rise to the situation indicated in the

following figure where Dialect C has Rule 2 followed by Rule 1 and Dialect D has the opposite order.

This hypothetical example is intended to suggest ways in which synchronic dialects may come to have rules identical but in different orders. (It is also possible that a rule is borrowed into different grammars at different positions in the grammars.) However, when we are dealing not with synchronic dialects but with different chronological stages of a language, there is no reasonable appeal we can make to some variant of the notion of wave. For concreteness let us take the two rules that were ordered oppositely in two distinct chronological stages of German. Middle High German had only Terminal Devoicing, the intermediate stage Early Modern German had Terminal Devoicing followed by Vowel Lengthening, and Modern Standard German has the two rules in reverse order. In order to develop an explanation in terms of different rates of diffusion of the two rules, similar to our hypothetical example, we would have to posit a hypothetical dialect whose grammar possessed Vowel Lengthening but not Terminal Devoicing. This hypothetical dialect would have to be assumed contemporaneous with Middle High German.

Now the stage is set for a wave explanation. Middle High German had Terminal Devoicing but not Vowel Lengthening; the hypothetical German dialect had Vowel Lengthening but not Terminal Devoicing. We assume different rates of diffusion from these two dialects. In one dialect or group of dialects the result is the order Terminal Devoicing followed by Vowel Lengthening (Early Modern German, some archaic modern dialects); in the other dialect area the result is Vowel Lengthening followed by Terminal Devoicing (Modern Standard German).

The catch in this is, of course, the hypothetical dialect which had only Vowel Lengthening but not Terminal Devoicing. There is not one single hint of evidence that such a dialect ever existed. There is no scribal evidence pointing to such a dialect. Even in those few German dialects or languages (like Yiddish) derived from German that today lack Terminal Devoicing, relic

forms point to the earlier existence of a rule devoicing some or all obstruents word-finally. (This was already noted under RULE LOSS.) And bear in mind that each time a case of rule reordering is presented, it will be necessary to assume pro forma the existence of some hypothetical dialect having one but not the other rule.

These exemplify the fatuous lengths to which one is led if rule reordering is to be ascribed to a wave effect. It is obvious that this is quite simply the wrong explanation, and that rule reordering with respect to two chronological stages of a language comes about through some different kind of mechanism. It will be suggested in Chapter 4 that this mechanism is simplification of a particular type, but this proposal will have to be deferred until language acquisition in the child has been discussed. At present we shall continue our enumeration of the categories of primary change.

SIMPLIFICATION. One of the most common ways in which dialects differ is in the generality of analogous rules in their grammars. Let us consider one rather simple example. As has been pointed out before, most German dialects have a rule that devoices final obstruents (Rule 3.11). In some dialects, however, the rule is less general: in Alsatian, for example, it affects only word-final fricatives (Becker 1967:112–113). This version of the terminal devoicing rule was stated already as Rule 3.16:

3.11   [+ obstruent] → [− voice] / ___ # (All obstruents affected)

3.16 $\begin{bmatrix} + \text{obstruent} \\ + \text{continuant} \end{bmatrix}$ → [− voice] / ___ # (Only fricatives affected)

Rule 3.11 is simpler: it has a feature count of three while Rule 3.16 has a feature count of four. (≠ is arbitrarily assigned here a value of one.) Rule 3.11 is also the more general of the two since it applies to the natural class of all obstruents whether stops or fricatives, and Rule 3.16 applies only to the natural class of fricatives. As regards the terminal devoicing rule, then, the difference between Alsatian and those dialects with Rule 3.11 is that the grammar of Alsatian has a less general, more restricted version of the rule. The lesser generality of Rule 3.16 is reflected formally in its higher number of features.

It seems probable that it is precisely in this way that dialects often differ. In a detailed generative phonological study of three modern German dialects, Becker (1967) found that their grammars differed most often in the increased generality, lessened generality, or presence of a given rule in one grammar but not the other. (In his study only one case of rule reordering was discovered.) We may cite here one of his examples of a typical situation. A given phonological rule whose structural change need not concern us affects /t/ in

Darmstadt Hessian, /p t/ in Alsatian, and /p t k/ in Züritütsch (the Swiss German dialect spoken in the city of Zürich). The structural analysis of this rule thus contains in the different grammars the segments:

/t/	/p t/	/p t k/
+ obstruent	+ obstruent	+ obstruent
− continuant	− continuant	− continuant
− voice	− voice	− voice
+ coronal	+ anterior	
+ anterior		

The decreasing feature counts of five, four, and three reflect increased generality (Becker 1967:59).

If as a common characteristic of their difference dialects display rules in different degrees of generality, we would expect to find that diachronic stages of the same language similarly differ in the generality of given rules. The earliest records of Old English indicate that the grammar contained a rule like Rule 3.16 affecting only word-final fricatives. Later records show final devoicing applying to all obstruents, not only fricatives; therefore later grammars of Old English contained the simpler (and more general) Rule 3.11.

Before proceeding, let us consider for a moment the relation between simplicity and generality in sound change and dialect borrowing. In the preceding examples a rule was judged more general if it applied to larger natural classes. The later terminal devoicing rule in Old English applies to both stops and spirants, whereas the earlier rule applied only to spirants. The later rule is simpler because it contains fewer features (three against four in the earlier rule): in the structural analysis the feature [+ continuant] is suppressed in passing from Rule 3.16 to Rule 3.11. Frequently, increased generality is expressed in just this way—by suppression of a feature. Yet it is clear that more is involved in assessing the generality of a rule (the naturalness of a rule) than simply the number of features. Consider Rule 3.17, which converts b to d, and Rule 3.17', which changes b to p:

3.17 $\begin{bmatrix} + \text{obstruent} \\ + \text{voice} \\ + \text{anterior} \\ - \text{continuant} \end{bmatrix}$ → [+ coronal]

3.17' $\begin{bmatrix} + \text{obstruent} \\ + \text{anterior} \\ - \text{coronal} \\ - \text{continuant} \end{bmatrix}$ → [− voice]

The two rules have the same feature count, yet intuitively one regards the change b > d as less natural, less expected than b > p. Our evaluation

procedure is incorrect in that it assigns to each rule the same degree of complexity, and there is no formal aspect of these rules which can be changed on a principled basis so that Rule 3.17' counts as more natural than Rule 3.17. Chomsky and Halle (1968:400-435) have proposed ways of remedying this serious defect in phonological theory. Here, having pointed out that generality is expressed in ways other than feature suppression, let us consider further cases of rule simplification.

In all examples previously discussed in this context, the environment has been the same in all variants of the rule: only the segment (or natural class of segments) has been different. An example of a different kind can be taken again from Becker (1967:68). All three dialects Alsatian, Darmstadt Hessian, and Züritüütsch have a rule that under certain conditions converts underlying /s/ to [š] within morphemes. One part of the environment can be stated as follows: *s* > *š* before [b d g] in Alsatian and Züritüütsch, *s* > *š* before [g] in Darmstadt Hessian. We have then two versions of a rule effecting the same change:

$$3.18 \quad \begin{bmatrix} + \text{obstruent} \\ + \text{continuant} \\ + \text{coronal} \end{bmatrix} \rightarrow [- \text{anterior}] \ / \ \underline{\quad} \begin{bmatrix} + \text{obstruent} \\ - \text{anterior} \end{bmatrix}$$

$$3.18' \quad \begin{bmatrix} + \text{obstruent} \\ + \text{continuant} \\ + \text{coronal} \end{bmatrix} \rightarrow [- \text{anterior}] \ / \ \underline{\quad} [+ \text{obstruent}]$$

Rule 3.18 is the rule in Darmstadt Hessian; Rule 3.18', in Alsatian and Züritüütsch.

The segment acted on by both variants of the rule is the same; only the environments are different. The environment of Rule 3.18' requires one feature. Rule 3.18' hence counts as more general, in agreement with our traditional assessment of generality.

One must not conclude from these examples that generality of rule is necessarily related to the number of segments affected by the rule. We do not consider a rule acting on /p z i r h/ more general than a rule acting on /u o ø/: the latter is a natural class, the former is not. "Generality" in phonology has traditionally been understood to involve "either more members of a natural class or more natural classes." The simplicity metric relates this notion to the number of features in the structural analysis of a rule, as the preceding examples illustrate.

An instance different in kind but not in principle is that of the umlaut rule in Continental West Germanic in contrast to the umlaut rule in Old English. In the grammars of the Continental West Germanic dialects (Old High

German, Old Saxon, and Old Frisian) the umlaut rule has the form of Rule 3.6:

$$3.6 \quad V \rightarrow [- \text{back}] \ / \ C\underline{\quad}_1 \begin{bmatrix} - \text{consonantal} \\ + \text{high} \\ - \text{back} \end{bmatrix}$$

This rule, which assimilates a vowel in backness to the [− back] high vowel or glide in the next following syllable, accounts for such alternations in Old High German as *gast* : *gesti* 'guest, guests', the latter from underlying /gast + i/; and *skōno* : *skōni* 'beautifully, beautiful', the latter from underlying /skōn + i/. Old High German orthography gives no indication of fronting in the ō of *skōni*, but we are assured of its presence by the later writing *schöne*, which does indicate fronting. In later Middle High German we find OHG *skōno* : *skōni* written *schöne* : *schœne*, and the reflexes in current German are *schōn* 'already' (no umlaut) and *schön* 'beautiful' (with umlaut). We thus assume that underlying /skōni/ in Old High German was realized phonetically as [skö:ni], though written without umlaut designation.

Old English, on the other hand, had a more general version of Rule 3.6. Its rule resembles Rule 3.6 in that back vowels were fronted before *i* and *j* in the next syllable (Old English has no long vowels under weak stress, so *ī* does not occur as an umlaut factor): *hlǣfdige* 'master, mistress'; *dōm* : *dōeman* (< *dōmjan*) 'judgment, to judge'; *brīcan* : *brȳčiþ* (< *brūciþ*) 'to need, he needs'; *gold* : *gylden* (< *guldin*) 'gold, golden'. But in addition to palatal umlaut (*i*-umlaut) Old English had in its earliest grammar a rule that backed front vowels before *ŭ* and *w* in the following syllable. Details in the formulation of this rule vary considerably from dialect to dialect (Brunner 1965:80-89, Campbell 1959:85-93); as examples we may cite *clipian* : *cliopode* (< *clipude*) 'call, he called'; *ġeset* : *ġeseoto* (< *ġesetu*) 'dwelling, dwellings'; *hæf* (< *haf*) : *heafu* (< *hafu*) 'sea, seas'. Velar umlaut of *i e a* produced *io eo ea*, which are assumed here to represent back unrounded vowels [ï ë a] in agreement with the views of Stockwell and Barritt (1955) and Hockett (1959). That there are no instances of *u*-umlaut of long vowels is due to various historical changes, one of which caused *u* to be deleted in pre-Old English after a heavy syllable—i.e. a syllable consisting of either (1) a long vowel plus any number of consonants including none, or (2) a short vowel followed by at least two consonants.

Thus, the umlaut rule in Old English—at least in the grammars of some of the earliest dialects—expressed a more general kind of assimilation than did Rule 3.6. The Old English rule not only assimilated back vowels to the frontness of following high vowels and semivowels, but also assimilated front vowels to the backness of following back high vowels and semivowels. For this we assume a rule:

3.19 $V \rightarrow [\alpha \text{ back}] / \underline{\hspace{1cm}} C_1 \begin{bmatrix} - \text{ consonantal} \\ + \text{ high} \\ \alpha \text{ back} \end{bmatrix}$

($\breve{u} \, \breve{o} \, \breve{a} > \breve{u} \, \breve{o} \, \breve{a}$ before at least one consonant followed by i or j; $i \, e \, a > i \, e \, a$ before at least one consonant followed by $\breve{u}$ or w. The backing rule did not apply to long vowels $i \, e \, a$ since these did not stand in the requisite environment. As mentioned before, the exact form of Rule 3.19 varied according to time and dialect, and no account of such differences is taken in its formulation here.)

Note that Rule 3.19 is a generalization of Rule 3.6 provided we observe the well-motivated convention that in the simplicity metric variables (such as alpha) count less than specified pluses or minuses. Rule 3.19 is clearly a generalization of Rule 3.6 since it effects not only the alternations given by Rule 3.6 but also the parallel set of changes involving the natural class of front vowels. The feature counts of both rules directly reflect this greater generality: they differ only in that Rule 3.19 has [α back] in place of a specified [− back] in the environment, and α counts less than a specified minus.

All examples offered so far have to do with the simplification (generalization) of grammar rules. Other kinds of simplification are relevant to linguistic change, some more trivial than rule simplification, some more subtle. Typically, in the history of a language, a lexical entry is simplified in ways that reduce the number of idiosyncratic features it carries. The change in the verb 'to help' between Middle English and Modern English is characteristic. In Middle English it was conjugated strong with the principal parts *helpe(n) halp hulpen holpen*; today it is weak *help helped helped*. What seems to have changed here is the set of features that determine this verb's morphological behavior. In Middle English it had schematically the lexical entry:

(help+ [+ Verb, + Strong, + Class III,...])

which characterizes it as a strong verb of the third ablaut class plus whatever else (...) is required to determine its grammatical behavior completely. In Modern English it has the lexical entry:

(help+ [+ Verb, − Strong,...])

which represents a simplification in that the single feature [− Strong] has replaced the two features [+ Strong, + Class III].

Instances of this kind are common, and the reason for describing them with the term "simplification" is obvious. It is perhaps less obvious that some of the types of primary change discussed earlier also represent an aspect of simplification. This is not in general true of rule addition in any obvious

---

sense, but it is true of rule loss and rule reordering. The simplification in the latter cases amounts to a reduction of allomorphic variation in certain morphemes at the surface level. Before the terminal devoicing rule was lost in Yiddish, a morpheme such as *veg* 'path' would have had two allomorphs: /vek/ and /veg/. The loss of the terminal devoicing rule in effect collapses these two allomorphs into one /veg/; so too for all the forms that display this type of biallomorphy.

In Standard German, before the terminal devoicing and vowel lengthening rules were reversed in order, a noun such as *Rad* 'wheel' would have had the two allomorphs /rat/ and /ra:d/, which differ both in the length of the root vowel and in the voicing value of the final obstruent. Upon reordering, *Rad* would have the allomorphs /ra:t/ and /ra:d/, which are different only in the voicing value of the obstruent.

The Monachi example presented in Section 3.2 as an instance of rule addition can be interpreted as a case of simplification. Recall that the Bishop dialect had nasalized [w̃] corresponding to non-nasalized [w] in the North Fork dialect. Nasalized glides are somewhat unusual in the world's languages; Chomsky and Halle (1968:407), in their marking conventions for glides, state that glides are specified non-nasal. Thus, a grammar that violates a universal constraint is more complex than one that obeys it. In the Monachi example we seem to have a transition from a grammar that violates a universal constraint on nasality in glides to a simpler grammar that obeys this constraint. Bishop represents the older, more complex stage; North Fork, the newer, simpler stage of grammar.

It is not altogether accidental that types of primary change turn out to be instances of simplification in disguise. Why this should be so is the problem to which we shall devote our attention in the following chapter.

SUPPLEMENTARY READING

Becker, Donald A. 1967. *Generative Phonology and Dialect Study: An Investigation of Three Modern German Dialects*. Unpublished doctoral dissertation, University of Texas at Austin.

Keyser, Samuel J. 1963. Review of Hans Kurath and Raven I. McDavid, Jr., *The Pronunciation of English in the Atlantic States*. *Language* 39.303-316.

Kiparsky, Paul. 1968. "Linguistic Universals and Linguistic Change," in *Universals in Linguistic Theory*, eds. Emmon Bach and Robert T. Harms. New York: Holt, Rinehart & Winston, Inc.

Saporta, Sol. 1965. "Ordered Rules, Dialect Differences, and Historical Processes." *Language* 41.218-224.

Weinreich, Uriel. 1954. "Is a Structural Dialectology Possible?" *Word* 10.388-400.