1968

Two Proposals Concerning the Simplicity Metric in Phonology

Emmon Bach
ebach@linguist.umass.edu

EMMON BACH, University of Texas

TWO PROPOSALS CONCERNING THE SIMPLICITY METRIC IN PHONOLOGY

1. A standard example for the influence of context on the variation of speech sounds is the determination of front and back allophones of the velar stops in languages like English. There must be few students of linguistics who have not been confronted in introductory courses with pairs like *keep* vs. *cool* or *peak* vs. *Luke*, *give* vs. *good* or *dig* vs. *dog*. A statement about this variation might take the form:

I. A velar stop varies in the dimension front to back (graveness) according to the quality of the preceding or following vowel.

How might this statement be rendered in a generative phonology of English, a theory which is intended to reflect generality of statement in shortness of rules? The best that we can do is the following.[1]

Rule 1.1

$$\begin{bmatrix} -cnt \\ +hi \end{bmatrix} \rightarrow [\alpha grv] \; / \; \left\{ \begin{bmatrix} -cns \\ \alpha grv \end{bmatrix} \; \underline{\quad} \;, \; \underline{\quad} \; \begin{bmatrix} -cns \\ \alpha grv \end{bmatrix} \right\}$$

1-The theory of phonology assumed here is essentially that developed by Morris Halle and associates, see Chomsky and Halle, 1968, and Harms, 1968. The features used in most of the examples follow modifications introduced by McCawley, 1967. Note in particular the use of the features *high* (*hi*): [+*hi*] replaces [+*dif*] for vowels, [-*dif*] ([+*cmp*]) for consonants: the features *obstruent* (*obst*), *syllabic* (*syl*), and *consonantal* (eliminating *vocalic*) for a four-way classification of segments: vowels are[+syl, -cns], (and [-obst]): glides are [-cns, -syl] (and [-obst]): liquids and nasals are [+cns, -obst]: stops, spirants, affricates are [+obst](and [+cns, -syl]). An earlier version of this paper was read at the Winter meeting of the Linguistic Society of America, December, 1966. I am indebted to Robert T. Harms for discussion and clarification of many of the problems raised in the paper.

It is apparent that we have here another Case of the Missing Generalization. For the simplicity metric, making use of current abbreviatory devices, would evaluate the above rule in exactly the same way as a rule we might write to express the following hypothetical statement:

II. A velar stop varies in graveness according to the graveness of a preceding vowel or the voicing of a following stop.

Rule 1.2

$$\begin{bmatrix} -cnt \\ +hi \end{bmatrix} \rightarrow [\alpha grv] \; / \; \left\{ \begin{bmatrix} -cns \\ \alpha grv \end{bmatrix} \; \_\_\_\_ \right\} \left\{ \begin{bmatrix} -cnt \\ \alpha vce \end{bmatrix} \right\}$$

I think it is obvious that I is a more general statement than II. Hence, we should be able to express I in a form that is evaluated more highly (given a lower measure by counting feature specifications) than II. My first proposal is, then, a new abbreviatory notation which can be applied in the first instance but not the second to express the greater generality of the first example and to capture the obvious fact that the same class of segments is specified in both parts of Rule 1.1. A rule like 1.1 will be abbreviated by eliminating the environment dash: a → b/c. Let us call the convention the 'neighborhood' convention. My arguments will be the following. First, I shall exhibit what I believe to be a significant number of examples of situations in different languages which can be expressed in a simple fashion only if we adopt the proposed notation (or some equivalent). Second, I shall present some historical evidence which supports the claim that the neighborhood convention expresses a 'significant generalization.'

2. Let us first state exactly what it is that we are trying to abbreviate by our convention. There are two points to consider. First, the order of the environments can be significant. Second, if we have an environment consisting of several segments, we must make some stipulation about the order of these several segments.

With regard to the first point, let us consider what happens to velar stops in English when *both* parts of the environmental specification can apply, that is, when we have velar stops that occur *between* vowels. If the two vowels are the same it will make no difference, of course. But consider examples in which the two vowels are different: *buggy, eager, Augie, bookie, Rocky, becalm*, etc. It is apparent that it is the second vowel which de-

termines the value of the stop. Thus, we tentatively decide that the abbreviation will be an abbreviation for the sequence of rules giving a value first *after* and then *before* a given environment, since whatever value is given to a segment which follows the determining segment, it will be changed by the second rule abbreviated by the convention.

Second, suppose we have a rule that involves an environment consisting of several segments, schematically '$a_1a_2...a_n$.' The question is this: Given an abbreviated rule '$c \rightarrow d \; / \; a_1a_2...a_n$ should this be defined as the abbreviation of Rule A or B?

$$\text{A. } c \rightarrow d \; / \; \left\{ \begin{array}{l} a_1a_2...a_n\text{——} \\ \text{——}a_1a_2...a_n \end{array} \right\} \quad \text{B. } c \rightarrow d / \left\{ \begin{array}{l} a_1a_2...a_n\text{——} \\ \text{——}a_n...a_2a_1 \end{array} \right\}$$

There seems to be fairly clear evidence that the second alternative is correct, that is, that the neighborhood abbreviation should be read 'after a certain environment or before the mirror-image of that environment.' I shall present below several cases where this decision works. I know of no rules involving a sequence of several segments where a particular specification is made after or before that sequence in the same order (which seems inherently implausible).

3. Let us now consider some examples of rules of this sort drawn from a number of languages. My reasoning here is that a notational convention can be partly justified by finding many examples of its use. It is true, as Kiparsky (1968) has argued, that there is no limit to the number of absurd notational conventions that could be invented and for which widespread 'use' could be found in the languages of the world. Thus, the value of examples such as those given here is dependent partly on the feeling of the linguist that what is expressed is somehow a valid generalization. Beyond that, however, the examples have a certain negative evidential value in that we would be suspicious of abbreviatory conventions for which we could not find fairly widespread employment.

3.1. The following statements are frequently made about the assimilations of voicing of stops and spirants in Modern Dutch

(e.g. Koolhoven 1962:19)[2]:

I. In a sequence of two stops, if one of the stops is voiced, so is the other.

II. In a sequence of a stop and a spirant, the stop determines the voicing value of the spirant.

III. A sequence of two spirants is always voiceless.

The regularities underlying these statements can be characterized by two ordered rules:

$$\text{Rule 3.1} \quad \begin{bmatrix} +\text{obst} \\ +\text{cnt} \end{bmatrix} \rightarrow [-\text{vce}] \; / \; [+\text{obst}]$$

$$\text{Rule 3.2} \quad [+\text{obst}] \rightarrow [+\text{vce}] \; / \; \begin{bmatrix} +\text{obst} \\ -\text{cnt} \\ +\text{vce} \end{bmatrix}$$

In the first rule all spirants are made voiceless in clusters with other obstruents. In the second rule all obstruents are voiced if they occur in the environment of a voiced stop. Clusters of spirants will not undergo Rule 3.2, hence will correctly remain voiceless.

Without the use of the neighborhood convention, it would be necessary to repeat the specifications in each environment. There would be no way to distinguish the degree of generalization achieved in the rules from that of two rules in which, say, spirants were made voiceless before obstruents or after glides and obstruents were voiced after voiced stops and before /r/. It should be noted that the order of expansion (i.e. the order of the abbreviated rules) can play no role in rules like these where a particular value is assigned to a feature. Hence, the crucial examples for settling the question of ordering will have to be found in rules involving Greek-letter variables (for assimilations and the like).

2-Aert Kuipers has pointed out to me that the statements made here about Dutch are oversimplified and perhaps incorrect. In his dialect, at least, there appears to be a terminal devoicing rule which operates within compounds and must take effect before the rules given above (or whatever turns out to be their correct formulation). Since I have seen the statements above in several published sources, I have let them stand, but it is likely that they must be modified at least to the extent of including some necessary restriction by means of various boundary elements.

3.2. In Modern German, clusters of obstruents are always voiceless within morphemes (and 'words' suitably specified by means of the appropriate boundaries--i.e. there can be sequences of voiceless and voiced obstruents only within compounds). Hence, whenever we have such a cluster it is not necessary to specify the feature of voicing in the underlying segments. We can add this specification by the rule.

Rule 3.3. $[+\text{obst}] \rightarrow [-\text{vce}] \; / \; [+\text{obst}]$

This rule states the fact noted above in the most direct possible way. Suppose we do not have some such device as the neighborhood convention. Using the usual formalism, we would have to repeat the environmental specification twice and again the rule would not reflect the obvious generalization. An alternative formulation might use the framework of a Structural Description and Structural Change:

Rule 3.4. SD: $[+\text{obst}]_1 \quad [+\text{obst}]_2$

SC: $1, 2 \rightarrow \quad [-\text{vce}]$

The defect of this notation is that there is no obvious way to show that it is more general than a rule with the same SD but a different SC, say '$1 \rightarrow [-\text{vce}]$', in which only one of the segments is affected. Since there are independent reasons for using the neighborhood abbreviation, there is no need to set up a special stipulation for evaluating rules like 3.4.

It should be noted that rules like those given above must be assumed to be embedded within a full analysis of the phonological system of the language in question (in so far as we can anticipate it) and must apply at the appropriate place in the sequence of rules. Thus, in German the phonetic obstruent [v] would seem to provide counter-evidence to the claim made since we have forms like *Zwei, quer* with [t^{s}v, kv] and the like. But there are various reasons for considering phonetic [v] in German as derived from an underlying semivowel /w/ (or perhaps even vowel /u/). The rule above is one small part of the justification for this analysis (Ross 1967, Bach and King, 1966). Further, anomalous forms like *Dschungel, Dschunk* will be marked as exceptions to the rule (note that in popular speech they usually have [tš]).

3.3. Examples like those cited in 1 (front and back allophones of velars) can be matched by many neighborhood assimilations

at the level of phonetic detail. Thus, in the immediate neighborhood of rounded (flat) vowels obstruents in English are usually more or less rounded. This is probably a universal rule. We may suspect such a situation to be the source for the Greek reflex of labiovelars as plain velars before or after /u/ (Buck 1933:126). Similarly, in Russian the allophones of certain vowels are determined by the presence of a palatalized consonant in the immediate environment (Halle 1959:73; Rules P 13a, b implicitly use the convention proposed here).

3.4. Let us consider next rules of vowel elision. A number of languages exhibit a series of rules for eliding vowels in such a way that the vowels can be set up in a hierarchy of 'dominance' such that more dominant vowels elide less dominant ones. For instance, in Modern Greek (Koutsoudas 1962:16-17) it seems that the following rule applies:

Rule 3.5

$$V \rightarrow null \; / \; \left\{ \begin{matrix} a \\ o \\ u \\ i \end{matrix} \right\} \; + \qquad$$

That is, any vowel is elided by a neighboring /a/ with intervening morpheme boundary (+), any vowel except /a/ by /o/, any vowel except /a/ and /o/ by /u/ etc. This rule is especially interesting since it shows the necessary order of expansion of two abbreviatory devices (this was pointed out to me by Stanley Peters). That is, if we expand the environment above first into

$$\left\{ \begin{matrix} a+ \\ o+ \\ u+ \\ i+ \end{matrix} \right\}$$

and then into

$$\left\{ \begin{matrix} a + \text{——} \\ \text{——} + a \\ o + \text{——} \\ \text{——} + o \\ \text{etc} \end{matrix} \right\}$$

we will get the right results, otherwise not. Also, Rule 3.5 provides our first example in support of the 'Mirror-image' character of our notation. Any such rule working across a boundary will be of this character and it is difficult to imagine a rule of the form

$$a \rightarrow b \; / \; \left\{ \begin{matrix} c + \text{——} \\ \text{——} + c \end{matrix} \right\}$$

dealing with such phenomena (e.g. a certain vowel is elided if it

follows a second vowel with intervening boundary or if it precedes the second vowel in prefinal position).

Similarly, in Amharic (extrapolating slightly from the statement in Obolensky et al., 1964:24) the three central vowels reduce when they occur together according to the hierarchy (from most to least dominant): a ə ɨ. We can give the following rule to account for these elisions:

Rule 3.6

$$[+\text{syl}] \rightarrow \textit{null} \;/\; \begin{bmatrix} +\text{syl} \\ \begin{Bmatrix} +\text{cmp} \\ -\text{hi} \end{Bmatrix} \\ +\text{hi} \end{bmatrix} \quad +$$

(Note the order of expansion: first the braces are placed around the segment brackets:

$$\left\{ \begin{bmatrix} +\text{syl} \\ +\text{cmp} \end{bmatrix} \begin{bmatrix} +\text{syl} \\ -\text{hi} \end{bmatrix} \begin{bmatrix} +\text{syl} \\ +\text{hi} \end{bmatrix} \right\} \quad +$$

then we proceed as above. I am assuming it is not necessary to specify [-tns] or whatever is used to distinguish the central vowels because of rules inserting glides between other vowels).[3]

3.5. A number of languages like German exhibit neutralization of voicing oppositions and the like before word boundaries. Sometimes, however, such neutralizations (as well as allophonic variations) are determined merely by the presence of a word boundary (hence, the common phrase in phonological descriptions 'initially and finally'). Thus, in Tübatulabal voiced stops and affricates occur only in 'intersonantic, syllable-initial position' Swadesh and Voegelin 1939), and we may state a rule:

Rule 3.7. $[-\text{cnt}] \rightarrow [-\text{vce}] \;/\; \left\{ \begin{matrix} \neq \\ [+\text{obst}] \end{matrix} \right\}$

Similarly, in Amharic geminate consonants are simplified initially and finally (but only after a rule lengthening vowels before single consonants.

I have by no means exhausted the number of examples of situations where neighborhood rules seem to offer the best method

3-I have drawn here upon work done on Amharic by Marvin Bender at the University of Texas at Austin.

of expressing the regularities apparent at various levels in phonological rules.[4] One particularly interesting set of facts are those of various kinds of 'vowel harmony,' but since such rules involve the second convention to be considered here, I shall defer discussion until a later part of this paper (7.).

4. So far my arguments in favor of the neighborhood convention have been of two sorts. On the one hand, they have been based on the 'linguist's intuition' that one rule is more general than another. And on the other hand, I have given a sampling of examples that show that the convention is a useful device. Let us now consider some stronger evidence in support of the idea that the convention does indeed capture 'significant generalizations.' This evidence is drawn from the history of German.

4.1. An important type of sound change is that in which a special rule becomes generalized. That such changes take place is a matter of historical fact. Moreover, such changes have been observed in the making. William Labov (1965) cites two cases. On Martha's Vineyard a centralization of the first element of /ay/ was followed by a centralization of /aw/, that is to say, the environment of a rule producing this variant of the compact vowel /a/ was simplified --i.e. generalized--by dropping the graveness specification:

4-See Becker 1967 for several instructive examples in three German dialects. In a paper read at the Winter meeting of the LSA, 1966, Kostas Kazazis discussed an interesting problem in Modern Greek involving the reduction of /i/ to a glide before and after another vowel, where in one layer of vocabulary the order of environments is as in the neighborhood convention posited here, in the other ('nonlearned' forms) it is just the opposite. This would seem to be an especially interesting case to consider in line with the question of historical evidence for the generality of such rules as considered in the next section. Note also that if we reverse the values for 'syl' in this case, that is, start with underlying /j/, then the order of environments will be just the opposite. James Fidelholtz (personal communication) has pointed out several cases in Micmac which can be neatly handled with the neighborhood convention, presumably to be discussed in his M.I.T. dissertation on Micmac phonology. It may be, incidentally, that when phonological theory is enriched to include more constraints on the substance of rules, it will turn out that particular orders of environment and affected segment will depend on the actual segments specified. Thus, nasals typically assimilate to following stops, but in most languages that I know of where stops can be followed by (nonsyllabic) nasals, the corresponding assimilation does not occur, e.g. German *Bank* ([ŋ k]) but *Knabe* ([kn]).

$$/ \_\_ \begin{bmatrix} -\text{cns} \\ -\text{syl} \\ -\text{grv} \end{bmatrix} \qquad / \_\_ \begin{bmatrix} -\text{cns} \\ -\text{syl} \end{bmatrix}$$

In this case, the simplicity metric (based on counting feature specifications) correctly reflects the greater generality of the second environment. Similarly, in New York (Labov 1965:106) raising of /eh/ (in Labov's transcription) was followed by raising of the corresponding nongrave vowel /oh/. For an example from history we may cite the change of *þ* (orthographic *th, dh, ð* etc.) to *d* in Old High German, which took place first medially and finally and then was generalized to all positions (Braune-Mitzka 1963: §§ 166-7).

A consideration of the way in which languages are learned and rules borrowed makes the process of rule generalization seem quite natural. Suppose--as in the cases given by Labov above--that there are two closely related and contiguous dialects A and B and that A is a more prestigious dialect in some sense. Suppose that dialect A undergoes a sound change of a very restricted sort, adding a late phonetic rule with a special environment, say, one lowering high front vowels if immediately followed by [n]. It may happen then that speakers of B under the influence of A will add a new rule to their grammar. On a priori grounds it seems likely that such a new rule will be at least as general as the rule of A, quite possible that it will be more general, highly unlikely that it will be more special. Thus, in the case considered speakers of B might add a rule lowering both /i/ and /u/ before [n], or the same before any nasal (any segment marked [+nas]). We would not expect them to add a rule with a more restricted environment such as 'before a cluster of [n] and a stop,' or 'between [b] and [n].' For to suppose the latter we would have to imagine that speakers of B heard just the accidental set of forms that would be covered by such a rule, and furthermore added a *rule* formulated to cover exactly the heard forms. (Cf. Harms 1966a and 1967 for discussion and examples of the process of rule generalization.)

These considerations point up what I believe to be the fallaciousness of the following kind of argument. A very general sound change takes place in a certain dialect. Contiguous dialects show (with increasing distance) more and more specialized instances of the same phenomena. The argument is then that the change originated in the area that shows the most general form

of the rule and spread from the area of 'strongest' effect gradually becoming attenuated as it spread (like ripples in a pond). Some examples of this are the High German sound shift, the monophthongization of *ai* and *au* in Old High German and Old Saxon, and the NHG diphthongization of high tense vowels. (For typical arguments of this sort we may take A. Bach 1956: §§ 59, 70). Thus in Old Saxon *ai* and *au* were monophthongized everywhere, but in the OHG area only in special environments. Assuming that there is a direct connection, to believe the traditional account we must imagine that speakers of Old High German conservatively borrowed rules in such a way as to apply them only before /x r w/ and /x r/ and dentals respectively! It seems to me obvious that quite the opposite direction of spread must have obtained.

Kiparsky (1968) has stated an almost diametrically opposed view according to which adults borrow rules in a more specialized form (while children generalize when learning their language--we might call this the Theory of Generative Children and Taxonomic Adults). As Becker (1967) has pointed out, Kiparsky's one example--NHG diphthongization--breaks down. Since we have two theories leading to opposite interpretations and factual claims, we must look to independent evidence to settle the question. The cases given by Labov seem to contradict Kiparsky's claim directly. In the historical examples we must consider documentary evidence. In the case of the German monophthongization there does not seem to be any clear evidence one way or the other,[5] while the High German sound shift would obviously demand a full-scale study. Most accounts are shot through and through with interpretations so that it is exceedingly difficult to evaluate the statements made in the handbooks. Since there are clear cases of rule generalization and no clear cases of specialization, I conclude that the former type of change is the expected mechanism and use the hypothesis as a necessary part of the

5-We have much fuller and older documentation for Old High German than for Old Saxon. It is possible to follow the change in OHG documents from about 765 (Braune-Mitzka 1963: §§ 43, 45). The OS *Heliand* is dated at about 830, the *Genesis* later but the manuscripts are considerably later and most of the smaller texts come from the tenth and eleventh centuries (Holthausen 1921: §§ 13-20). As with the HG sound shift I can do no more than raise questions here that must be answered if possible by detailed philological work.

argument below. (To say that rules can be both generalized and specialized without adding some further factors determining one or the other effect is, of course, to say exactly nothing. For some discussion of this whole problem see Becker 1967).

4.2. If the above reasoning is correct, then various sound changes that can be expressed as differences in rules can be used as evidence for the hypothesis that the neighborhood convention expresses a significant generalization.

In the OHG area, the following sound changes took place in the order given (over a period of about 200 years, the last two approximately contemporaneous):

A. ai > ei
B. au > ou
C. ua > uo (< $\bar{o}$)
D. ia > ie (< $\bar{e}_2$)

It is crucial to the following argument to show that these sound changes have their counterparts throughout the period as actual synchronic rules in the phonology of the language. The full justification for this cannot be presented here but only briefly sketched. The evidence for the first two involves a consideration of the system of rules for the forms of strong verbs, which in Old High German (as in all the older Germanic dialects) are still completely phonologically predictable. The assumption of underlying diphthongs of the form *au*, *ai* (or possibly *aw*, *ay*) can be shown to yield a simpler set of rules for the preterit and past participle forms, since we have complete parallelism throughout for the first three classes (and partial parallelism for all of the first five classes--note that $\bar{i}$ is derived from underlying *ei*):

I	reit-	rait	rit-	
II	fleug-	flaug	flug-	(-flog-)
III	helf-	half	hulf-	(-holf-)
IV	stel-	stal		(-stol-)
V	geb-	gab		

With some simple rules of adjustment of the sort considered here all the strong verbs can be handled with two or three rules. The total set of rules then can be shown to be much simpler than would be the rules which did not incorporate the synchronic correspondents to A and B above. Further support comes from the rules for *a*/*e* umlaut. The *ua* and *ia* are not underlying forms but intermediate between systematic phonemic and phonetic. The justification of rules corresponding to C and D is rooted in the treatment of the socalled 'reduplicating' verbs (which again can

be made quite simple).

Now, the series of changes given above can be shown to be just a progressive generalization of a single rule:

Rule 4.1 $\begin{bmatrix}+syl\\+cmp\end{bmatrix} \rightarrow \begin{bmatrix}-cmp\\-grv\end{bmatrix} / \_\_ \begin{bmatrix}-cns\\+hi\\-grv\end{bmatrix}$

Rule 4.2 $\begin{bmatrix}+syl\\+cmp\end{bmatrix} \rightarrow \begin{bmatrix}-cmp\\\alpha grv\end{bmatrix} / \_\_ \begin{bmatrix}-cns\\+hi\\\alpha grv\end{bmatrix}$

Rule 4.3 $\begin{bmatrix}+syl\\+cmp\end{bmatrix} \rightarrow \begin{bmatrix}-cmp\\\alpha grv\end{bmatrix} / \begin{bmatrix}-cns\\+hi\\\alpha grv\end{bmatrix}$

Rule 4.2 is more general than 4.1 if we incorporate some way of counting Greek letter variables as less costly than pluses or minuses. But 4.3 can be counted at least as general as 4.2 only if we adopt the neighborhood convention. Furthermore it can be evaluated as more general than 4.2 only if we stipulate in some way that the environment dash '——' is to be counted in such a rule as 4.2.

5. The second situation I would like to consider is that in which the determining environment for some rule is separated from the segment undergoing the rule by an irrelevant stretch of segments. Consider these two statements:

I. A mid front vowel becomes high front when immediately followed by a high front vowel.

II. A mid front vowel becomes high front when followed by a high front vowel.

Since Statement II includes Statement I as a special case it would appear to be more general. Yet as rules we have the following:

Rule 5.1 $\begin{bmatrix}+syl\\-grv\\-cmp\end{bmatrix} \rightarrow [+hi] / \_\_ \begin{bmatrix}+syl\\-grv\\+hi\end{bmatrix}$

Rule 5.2 $\begin{bmatrix}+syl\\-grv\\-cmp\end{bmatrix} \rightarrow [+hi] / \_\_ X \begin{bmatrix}+syl\\-grv\\+hi\end{bmatrix}$

I am not aware of any discussion in print of the role of variables like 'X' in the simplicity metric (but see Harms 1968). If we choose to count them then 5.2 will be evaluated as less general than 5.1. If we do not, then the two rules will be equally simple. In either case the greater generality of 5.2 is not expressed by the metric.

Let us consider, then, that a rule like 5.2 is to be counted as simpler than a corresponding rule without the variable X. A discussion of the exact details of this evaluation seems somewhat academic at the moment, in view of the many problems involved in deciding how to count symbols like 'C_o', 'C_1^2' and the like. Since 'X' is defined to be any sequence of segments '$a_1\ a_2\ \ldots\ a_n$' with $n \geq o$, we may think of the measure as stipulating that we count in some manner the concatenation sign which is 'officially' present in the Rule 5.1:

$$/\ \_\_\ \frown \begin{bmatrix} +syl \\ -grv \\ +hi \end{bmatrix}$$

but covered by the variable in 5.2.

Note that the notion of having rules of various 'levels' (depending on the hierarchy of boundary symbols, that is, morpheme-level, word-level rules, etc.), makes it unnecessary to add any special stipulations about the boundary symbols that may be covered by 'X'. Further, a rule like 5.2 makes no sense without such a hierarchy of levels. If 5.2 is a word-level rule then the stretch covered by 'X' may contain any boundaries lower than word boundaries, if phrase-level, then any boundaries lower than phrase level and so on (on the hierarchy of boundary-symbols see McCawley 1965, also Harms 1968).

5.1 Since I am not introducing a new notational convention here, but rather suggesting a method of evaluating rules involving an already widely used device, it will be unnecessary to adduce examples in any great number. The most obvious use for variables like 'X' is, perhaps, in rules for vowel-harmony and the like, where a value of some segment determines the value of all the other segments of a certain type in a 'word.' The two rules for vowel harmony in Turkish, for example, differ precisely in that 'graveness-harmony' involves an X-variable, while 'flatness-harmony' does not:

Rule 5.3 $[+syl] \rightarrow [\alpha grv] \ / \begin{bmatrix} +syl \\ \alpha grv \end{bmatrix} \ X \ \_\_$

Rule 5.4 $\begin{bmatrix} +syl \\ +hi \end{bmatrix} \rightarrow [\alpha flt] / \begin{bmatrix} +syl \\ \alpha flt \end{bmatrix} C_o \left(\begin{bmatrix} +syl \\ +hi \end{bmatrix} \ C_o \right) \begin{matrix} o \\ o \end{matrix}$

Thus, two forms like *geldikleri* 'their having come' and *güldükleri* 'their having laughed' (Swift 1963:46) harmonize throughout in graveness, but the flatness of a sequence of high vowels is de-

termined by the immediately preceding vowel (if not, that is if we rewrote 5.4 with 'X' we would get **güldüklerü*). Other examples for X-rules are rules for stress assignment, consonant-harmony and the like. I return to the question of vowel harmony below (7).

6. Once again we may turn to history for evidence that a rule involving an 'X' is more general than a rule without such a variable.

The regular correspondent for IE *ei* (or ey) in Germanic is *ī*. An analysis of the Ablaut system of the older dialects as alluded to above together with a consideration of the vowel systems of these dialects shows that we may posit a rule for Proto-Germanic of the following form:

Rule 6.1 $\begin{bmatrix}+\text{syl}\\-\text{cmp}\\-\text{grv}\end{bmatrix} \rightarrow [+\text{hi}] \;/\; \_\_ \begin{bmatrix}-\text{cns}\\+\text{hi}\\-\text{grv}\end{bmatrix}$

This rule remains synchronically operative in the older dialects for essentially the same reasons as those sketched above (4.2).

Apparently later this rule was generalized by the insertion of 'X':

Rule 6.2 $\begin{bmatrix}+\text{syl}\\-\text{cmp}\\-\text{grv}\end{bmatrix} \rightarrow [+\text{hi}] \;/\; \_\_\text{X} \begin{bmatrix}-\text{cns}\\+\text{hi}\\-\text{grv}\end{bmatrix}$

(The chronology is quite unclear, cf. Hirt 1931, Krause 1963. It should be noted that the generalization here can work in either direction, by inserting 'X' as above where the rule *ei* ⇒ *ī* now applies when the determining element is in another syllable, or changing 'C_o' to 'X' in which case the rule now applies to immediately contiguous segments). The rule now accounts for alternations like OHG *helfan/ hilfis*. In Old High German (and Old Saxon) the rule is further generalized by striking the specification [-grv] from the environment, so that we also get *hilfu* etc.

7. I now turn to a number of interesting cases in which both of the conventions discussed above can play a role. The first is taken from Diola-Fogny (Sapir 1965). Fogny has a set of five lax and five tense vowels. In any word the vowels are all either lax or tense, but the situation is not exactly a typical case of vowel harmony (as Sapir calls it) since tense vowels are 'dominant'. That is, rather than having a rule with alpha-variables (as in

Turkish or Finnish) we have a rule which assimilates lax vowels to tense vowels. The assimilation works in both directions and is independent of the distinction between root and affix. We might state the rule as follows:

$$\text{Rule 7.1} \quad [+\text{syl}] \rightarrow [+\text{tns}] \; / \; \begin{bmatrix} +\text{syl} \\ +\text{tns} \end{bmatrix} \; \underline{\quad X \quad}$$

A word consisting of morphemes with inherently lax vowels will (correctly) escape the effect of the rule. If any tense vowel occurs in a word then all the vowels will be tense. If all vowels are inherently tense then the rule applies vacuously. Thus, it seems as if we have satisfactorily accounted for the facts presented by Sapir (11-12).

The rule above works correctly, it will be noted, only if we assume that all vowels are specified for tenseness at the point where the rule applies. Since there are no morphemes containing both tense and lax vowels it is possible to specify the tenseness value for just one vowel (say the first) and predict the tenseness of following vowels by means of a morpheme-structure (MS) rule:

$$\text{Rule 7.2} \quad [+\text{syl}] \rightarrow [\alpha\text{tns}] \; / \; \begin{bmatrix} +\text{syl} \\ \alpha\text{tns} \end{bmatrix} \; \underline{X \quad}$$

Actually, both 7.1 and 7.2 can be simplified to cover all segments, cf. Sapir 1965:5: 'all consonants are tense in the presence of tense vowels.' Diola-Fogny then provides another case of allophonic and phonemic variation determined by the same rule.[6]

6-In a theory of phonology which restricts blank-filling rules to the MS component it is necessary to have a separate rule like 7.2, even for the true vowel-harmony situations considered below. It might be objected that essentially the same constraint is being handled in two different places in the grammar. If all vowels in words must agree in tenseness, then a fortiori so must they within morphemes. But within current phonological theory it is impossible to use one rule both to fill in redundant specifications and switch features in the case of Diola-Fogny. Note, however, that if the theory is modified to include an asymmetry of feature specifications in terms of marked versus unmarked values, then a rule like 7.1 can be used to give the marked value for tenseness to all segments in the presence of one marked-tense segment within the word, while presumably universal rules would apply to interpret unmarked-tense vowels as lax, marked-tense vowels as [+tns]. Sapir says that the tense vowels are relatively higher and 'closer than the corresponding lax vowels'. Whether the distinction involves merely height or muscular tension or tongue-root retraction as in many African vowel-harmony systems, it would seem reasonable to consider the tense vowels marked. If the simplicity measure then counts only marked values in the lexicon, we find a corroboration in the frequencies in the lexicon: lax vowels appear to be about ten times as frequent as tense vowels in Diola-Fogny (estimate based on Sapir's morpheme index and a short text on p. 116).

7.1. Two writers have dealt recently with the general problem of vowel harmony within the framework of generative phonology (Lightner 1965: Zimmer 1967). Lightner, using Classical Mongolian as a basis, has argued in favor of using morpheme features like GRAVE (not to be confused with the phonological feature). His arguments against a more traditional analysis (based on phonological features and involving rules determining gravity from left to right) may be summarized as follows: (1) an arbitrary decision must be made as to whether to specify the first or last vowel in a root as to its gravity, since (2) velar obstruents agree in gravity with preceding and following vowels and (3) if one has to have both progressive and regressive rules for consonants, why not also for vowels? Since by a general convention (i.e. a universal of language) every segment in a word is specified for every morpheme feature of the root,[7] it is possible to state a very simple rule of roughly the form

$$\text{Rule 7.3} \quad \left\{ \begin{bmatrix} \text{+syl} \end{bmatrix}, \begin{bmatrix} \text{+obst} \\ \text{-str} \\ \text{+hi} \end{bmatrix} \right\} \rightarrow [\alpha\text{grv}] \; / \; \begin{bmatrix} \text{——} \\ \alpha\text{GRAVE} \end{bmatrix}$$

(adapting Lightner's rule to the feature system used in the present paper).

The introduction of the neighborhood convention undercuts one of the arguments in favor of Lightner's solution. As Zimmer points out, it is unlikely that the determination of the quality of the velar consonant is brought about by the same rule that effects the vowel harmony. But if it is, then the rule for choosing the correct velar segment can be given as a neighborhood rule very similar to the rule for choice of velar allophones in English with which we began our discussion (1). To Zimmer's arguments one can add

7-As it stands Lightner's statement cannot be accepted ('each phonological segment of a *word* is associated with the abstract markers of the *root*' (248, fn. 9, my emphases). Lightner's arguments would apply to the case of Diola-Fogny considered above but the abstract marker TENSE that he would presumably posit is not determined for every segment of a word by the value for the root. There are further many cases in many languages where root morphemes that carry abstract features like +FOREIGN, +LEARNED are combined with perfectly regular (-FOREIGN, -LEARNED etc.) affixes.

that such a rule is probably a universal rule that does not need to be stated in the phonology of Classical Mongolian. But even if we accept Lightner's formulation we can maintain the uniform assignment of the gravity value to the first vowel of the root and restate 7.3 as follows:

Rule 7.4 $\left\{ \begin{matrix} [+syl] \\ \\ \begin{bmatrix} +obst \\ +hi \\ +str \end{bmatrix} \end{matrix} \right\} \rightarrow [\alpha grv] \; / \; \begin{bmatrix} +syl \\ \alpha grv \end{bmatrix} \; X$

By the conventions assumed above we expand this rule first into

Rule 7.4' A. $[+syl] \rightarrow [\alpha grv] \; / \; \begin{bmatrix} +syl \\ \alpha grv \end{bmatrix} \; X$

B. $\begin{bmatrix} +obst \\ +hi \\ -str \end{bmatrix} \rightarrow [\alpha grv] \; / \; \begin{bmatrix} +syl \\ \alpha grv \end{bmatrix} \; X$

Then into

Rule 7.4'' A. $[+syl] \rightarrow [\alpha grv] \; / \; \left\{ \begin{matrix} \begin{bmatrix} +syl \\ \alpha grv \end{bmatrix} \; X\_\_ \\ \\ \_\_X \; \begin{bmatrix} +syl \\ \alpha grv \end{bmatrix} \end{matrix} \right\}$

B. $\begin{bmatrix} +obst \\ +hi \\ -str \end{bmatrix} \rightarrow [\alpha grv] \; / \; \left\{ \begin{matrix} \begin{bmatrix} +syl \\ \alpha grv \end{bmatrix} \; X\_\_ \\ \\ \_\_X \; \begin{bmatrix} +syl \\ \alpha grv \end{bmatrix} \end{matrix} \right\}$

The convention for expanding the remaining abbreviations is that rules with 'X' apply first to the longest domain possible, then the next longest etc. (Harms 1968). Now it can be easily seen that 7.4 operates correctly. First all vowels in a word are made to agree in gravity with the first vowel (necessarily the first vowel of the root in Classical Mongolian). The second part of 7.4'' A applies vacuously. Now, the rule (7.4'' B) applies to bring all velars into line with the vowels whether they follow or precede

the vowels.

Zimmer (1967) has presented several arguments against Lightner's use of an abstract root marker GRAVE for languages like Classical Mongolian, Finnish and Turkish (although at the end of his note he seems to retract that position). There is no need to repeat his discussion here. It seems to me, however, that there is one cogent argument against a treatment like that of Lightner's which has not been put forward. Rule 7.3 looks like a very straightforward and simple rule but its generality is spurious. Note that the feature GRAVE is a completely ad hoc feature which has nothing to do inherently with the phonological feature 'grave'. It bears no more relation to the phonological feature of gravity than would, say, the features NOUN or ANIMATE. Moreover, unlike the latter features it plays no role in the language outside of just this rule. If we consider that such ad hoc features are allowable there is nothing to prevent us from effecting the following 'economy' in a language like German or English. In German there is, of course, no constraint on the kinds of vowels (say, in terms of gravity) which can occur in a word or root. Nevertheless, a significant portion of the lexicon will consist of items which exhibit only grave or non-grave vowels. Suppose we set up the abstract markers HARMONIC and GRAVE. Now we can leave out the specification of the phonological feature of 'grave' for all items which are marked +HARMONIC and reintroduce these 'predictable' specifications by a rule very much like 7.3 (applying to words like *Mákulatur, Uhu, Esel, Auto, Kuckuck, Popo, Vopo, Natur,* etc.: note that we can exploit various further 'generalizations': all monosyllables are predictably +HARMONIC: if a form consists of two syllables and is -HARMONIC, then the gravity of the second vowel can be predicted from that of the first, and so on).

7.2. One morpheme feature which is typically relevant to situations of vowel harmony is one which is not ad hoc like GRAVE but rather independently justified, namely the feature which distinguishes roots or stems from affixes. It seems that such a distinction often plays a role throughout a phonological system, especially in determining stress assignment and the like, but also in determining segments. Thus, in German, inflectional suffixes are built out of a very small subset of the segments of the language; moreover, the selection of the segments seems to exploit the maximal differentiations in the system: one stop /t/, one spirant /s/, one liquid /r/, one vowel /ə/, one nasal /n/ (with

minor exceptions). Further, it seems likely that we can define this distinction in a quite natural way in terms of the machinery of a grammar: by and large, lexical elements will be -AFFIX, everything else +AFFIX. That is to say, it is not necessary to mark elements for this feature (as it is for GRAVE), since it is predictable in a language-independent way.

Zimmer adduces Igbo as an example of a language in which both suffixes and prefixes harmonize.[8] In such languages one can give a straightforward rule corresponding to the statement: the tenseness of the vowel in an affix is determined by the tenseness of the vowel of the root. Zimmer takes such situations as precisely those in which we would use Lightner's solution by use of a morpheme feature like TENSE, but with the feature AFFIX it is not necessary to use any such ad hoc device:

Rule 7.5

$$[+\text{syl}] \rightarrow [\alpha\text{tns}] \; / \; \begin{bmatrix} +\text{syl} \\ \alpha\text{tns} \\ -\text{AFFIX} \end{bmatrix} \; X$$

Note that the adoption of the various conventions used here--the neighborhood convention for after and before (in that order), the expansion of rules with 'X' to apply first to the longest domain--and the principle that rules must be stated in the most general possible way lead us to a decision that it must be the first vowel of a stem or root that is distinctively marked for the feature in question: graveness in Finnish, Turkish, Mongolian, tenseness in Igbo, Ijo (Williamson 1965). We know of languages where this is the correct decision, namely those in which the rules for

8-There seem to be a number of such languages in West Africa. Jakobson, Fant, Halle (1951:59-60) also mention Bari and Maasai--incidentally the reference given (p. 61) to Tucker and Mpaayei 1955 is wrong: p. 260 is a page in the glossary, apparently what is meant is 240 but see the entry in the index under 'vowel harmony' for a number of cases where affixes determine the quality of vowels in the root. I suspect that the system is of the kind considered above for Fogny.

determining vowels of suffixes can also be used to determine the value for vowels within the root.[9] We know of languages where there are no independent grounds for choosing the first or the last vowel to be distinctively specified for some feature elsewhere redundant. In such languages, there is no reason to follow our conventions, but there is likewise no reason not to. I know of no languages where the choice forced on us by the conventions suggested leads to incorrect results. Such a language would be one, say, in which there was vowel harmony within the root and also (and only) harmonizing prefixes which would lead us to set up the last vowel as distinctive. The use of the neighborhood convention for situations like vowel harmony, in short, expresses the view that (at least in situations like vowel-harmony) the direction is predominantly progressive, possibly progressive and regressive (when one value is dominant like in Fogny or, apparently, Nez Perce, Aoki: 1966: or when lexical elements determine affix elements as in Igbo, Ijo, etc.) but never purely regressive. I take this as being strong evidence in favor of the linguistic reality of the proposed convention.

9-The arguments here depend in part on the assumption that a phonological rule can be used both to fill in redundant specifications within morphemes and to effect alternations in affixes and the like. In fact, all linguists that I know of who have dealt with vowel-harmony in a generative framework have made this assumption (Lees 1961: McCawley in several unpublished treatments of Finnish: Harms 1966b: Zimmer 1967: Lightner 1965). Without this assumption no morphophonemic evidence can be used to decide whether to mark the first or last segment of a morpheme for the feature in question. The adoption of the notation still forces us to favor progressive assimilations over regressive ones even within morphemes. It is difficult to see what synchronic evidence could be used to support this idea. It may be possible to use historical evidence (predominance of progressive assimilations ?), although it seems again (cf. fn. 4, end) that something more than a purely formal device is involved.

Aoki, Haruo. 1966. 'Nez Perce vowel harmony and Proto-Sahaptian vowels.' *Lg.* 42. 759-767.

Bach, Adolf. 1956. *Geschichte der deutschen Sprache.* 6. Auflage. Heidelberg.

Bach, Emmon and Robert D. King. 1966. 'Generative phonology of Modern Standard German.' Paper presented to the South Central Modern Language Association.

Becker, Donald A. 1967. *Generative phonology and dialect study: an investigation of three Modern German dialects.* Unpublished dissertation. University of Texas at Austin.

Braune, Wilhelm. 1963. *Althochdeutsche Grammatik.* Fortgeführt von Karl Helm, 11. Auflage bearbeitet von Walther Mitzka. Tübingen.

Buck, Carl Darling. 1933. *Comparative grammar of Greek and Latin.* Chicago and London.

Chomsky, Noam and Morris Halle. 1968. *The sound pattern of English.* New York, Evanston, London.

Halle, Morris. 1959. *The sound pattern of Russian.* The Hague.

Harms, Robert T. 1966a. *Review of Itkonen: Proto-Finnic final consonants. Lg.* 42. 825-831.

----------. 1966b. 'The measurement of phonological economy.' *Lg.* 42. 602-611.

----------. 1967. 'Split, shift and merger in the Permic vowels.' In *Ural-Altaische Jahrbücher.*

----------, 1968. *Introduction to Phonological Theory.* Englewood Cliffs.

Hirt, Hermann. 1931. *Handbuch des Urgermanischen I.* Heidelberg.

Holthausen, F. 1921. *Altsächsisches Elementarbuch.* Heidelberg.

Jakobson, Roman, G. G. M. Fant and Morris Halle. 1963. *Preliminaries to speech analysis.* Fourth printing. Cambridge, Mass.

Kazazis, Kostas. 1966. 'On the phonology of palatalized consonantism in Modern Greek.' Paper presented at the Winter meeting of the LSA.

Kiparsky, Paul. 1968. 'Linguistic universals and linguistic change.' In *Universals in linguistic theory,* eds. Emmon Bach and Robert T. Harms.

Koolhoven, H. 1962. *Teach yourself Dutch.* New York and London.

Koutsoudas, Andreas. 1962. *Verb morphology of Modern Greek: a descriptive analysis.* Bloomington, Indiana. (=Publication 24 of the Indiana University Research Center in Anthropology,

Folklore, and Linguistics.)
Krause, Wolfgang. 1963. *Handbuch des Gotischen*. 2. Auflage. Munich.
Labov, William. 1965. 'On the mechanism of linguistic change.' In Report of the Sixteenth Annual Round Table Meeting on Linguistics and Language Studies. Pp. 91-114. Washington, D.C.
Lees, Robert B. 1961. *The phonology of Modern Standard Turkish*. Bloomington, Indiana, and The Hague. (=Indiana University Publications, Uralic and Altaic Series, Vol. 6.)
Lightner, Theodore, M. 1965. 'On the description of vowel and consonant harmony.' *Word* 21.244-250.
McCawley, James D. 1965. *The accentual system of Standard Japanese*. Unpublished dissertation. Massachusetts Institute of Technology.
----------. 1967. 'The role of a phonological feature system in a theory of language.' *Langages*, No. 8. 112-123.
Obolensky, Serge, Debebow Zelelie and Mulugeta Andualem. 1964. *Amharic basic course*. Units 1-50. Washington, D. C.
Ross, John R. 1967. 'Der Ablaut bei den deutschen starken Verben.' *Studia Grammatica 6.47.118*.
Sapir, J. David. 1965. *A grammar of Diola-Fogny*. Cambridge. (=West African Language Monographs 3.)
Swadesh, Morris and Charles F. Voegelin. 1939. 'A problem in phonological alternation.' *Lg.* 15. 1-10.
Swift, Lloyd B. 1963. *A reference grammar of Modern Turkish*. Bloomington, Indiana, and The Hague. (=Indiana University Publications, Uralic and Altaic Series, Vol. 6.)
Tucker, A. N. and J. Tompo Ole Mpaayei. 1955. *A Maasai grammar*. London, New York, Toronto.
Williamson, Kay. 1965. *A grammar of the Kolokuma dialect of Ijọ*. Cambridge. (=West African Language Monographs 2.)
Zimmer, Karl E. 1967. 'A note on vowel harmony.' *IJAL* 33. 166-171.