Anderson (1974)

The Organization of Phonology

CHAPTER 13

How to Apply a Rule to a Form

In the chapters of Parts II and III, we have assumed that it was clear what was meant by 'applying' a rule to a given form. Our concern up to this point has been with the interaction of different rules, or with the same rule as it applies at different stages in a derivation, and not with the application of a single rule to a single form. Indeed, while there are problems in formalizing this notion, the vast majority of cases are entirely unproblematic. Given a string which contains an instance of the element A in the proper environment, say $WXAYZ$, it is clear what should be done to effect an application of a rule such as '$A \rightarrow B \mid X \_\_\_ Y$' to this string, giving as output '$WXBYZ$'.

Problems of a substantive nature begin to arise, however, when we consider the application of a rule to a string which contains several instances of substrings satisfying the rule's structural description. When several potential applications of a rule exist in a single string, how is the change specified by the rule to be carried out? Numerous alternatives can be imagined: apply the rule simultaneously everywhere; apply the rule in one place at a time, in random sequence; apply the rule in only one place, making the choice either randomly or on some universal or language particular basis; etc. These alternatives of course have concrete empirical consequences for the analyses of specific languages, and it is necessary to obtain evidence in order to decide among them. The relevant cases are by no means easy to find, however, and it is not surprising that this issue has only arisen in rather recent work.

Chomsky and Halle (1968) saw the existence of this issue, though they had no cases at hand to allow them to make a principled decision. Accordingly, the algorithm for rule application which they propose was intended simply to make as strong as possible a hypothesis about linguistic structure. Their choice was the simultaneous application of a rule at every possible place in the string:

> To apply a rule, the entire string is scanned for segments that satisfy the environmental constraints of the rule. After all such segments have been identified in the string, the changes required by the rule are applied simultaneously (Chomsky & Halle, 1968, p. 344).

This principle makes the claim that information about the application of the rule at one point in the string cannot be relevant to the application of the same rule at another point in the same string. It thus minimizes the extent of the interaction between different applications of the same rule. In order to maintain this principle completely, Chomsky and Halle were led to propose the device of simultaneously applied infinite schemata of rules [the ( )* notation] for the description of processes with potentially unlimited scope. In Chapter 9, we examined this device and found it to lead to the loss of generalizations in many cases. We proposed there that some rules should be allowed to reapply to their own outputs, thus introducing a certain amount of interaction among applications of the same rule in the same derivation.

Interactions of the sort discussed in Chapter 9, however, do not exhaust the range of cases that must be considered. Cases in which one application creates other potential applications are allowed in this way, but no decision has yet been reached on the matter of what to do when several potential applications exist in the same string. To make the problem concrete, consider a hypothetical language with a rule which applies stress to a vowel preceding and following an unstressed vowel:

$$(1) \quad [+\text{syll}] \longrightarrow [+\text{stress}] \,/\, \begin{bmatrix} +\text{syll} \\ -\text{stress} \end{bmatrix} C_0 \_\_ C_0 \begin{bmatrix} +\text{syll} \\ -\text{stress} \end{bmatrix}$$

Now consider the way this rule might apply to a form such as /CVCVCVCVCVCV/. We might choose to apply (1) in accord with the simultaneous principle suggested by Chomsky and Halle, in which case we would stress all but the first and the last of the vowels in this form. If, on the other hand, we should choose to apply the rule in one place at a time, proceeding from left to right, we would obtain an alternating pattern of stress, with the second and fourth but not the first, third, fifth, or sixth vowel stressed. This is because the application of the rule to the second vowel (the leftmost vowel that meets its structural description) removes the environment for its application to the third vowel, and the next vowel that can be affected is the fourth vowel. After the rule applies to the fourth vowel, it can no longer apply to either the fifth or the sixth. If we proceeded one step at a time from right to left in the same form, however, we would again obtain two stressed vowels, but this time these would be the third and the fifth.

If we can find cases, then, in which the application of a rule is possible in several different places, but in which some of these applications would have the effect of removing the environment for others, we can confirm or deny the predictions made by Chomsky and Halle. A case of this sort is discussed by Dell (1970), who deals with the rules affecting the vowel ə in French. Dell demonstrates the existence of a number of rules producing and deleting this vowel, including the two rules abbreviated by the schema (2):

$$(2) \quad \text{ə} \longrightarrow \emptyset \,/\, [+\text{syll}]\ (\#)\ C \_\_$$

This rule, which deletes schwa if it is preceded by just one consonant which is in turn preceded by a vowel, is optional in its application across word boundaries, though obligatory within words. Given a form such as *tu devenais* (/tü#dəvənɛ/) the rule could potentially apply to either of the two schwas. Application to the first schwa would destroy the context for the application of the rule to the second schwa. In fact, this form has two alternative pronunciations: either [tüdvənɛ], or [tüdəvnɛ], but not *[tüdvnɛ], with both schwas elided. In this case, a principle exists which allows us to obtain the correct results without regard to the application algorithm employed. Rule (2), as we noted, is an abbreviation for the two rules in (3):

$$(3) \quad \text{a.} \quad \text{ə} \longrightarrow \emptyset \,/\, [+\text{syll}]\ \#\ C \_\_$$

$$\text{b.} \quad \text{ə} \longrightarrow \emptyset \,/\, [+\text{syll}]\ C \_\_$$

It is only (3a) which is applicable to the first schwa in *tu devenais*, while it is only (3b) which is applicable to the second. Rule (3a) is optional, and precedes (3b); if we choose to apply (3a), it makes (3b) inapplicable, and we get [tüdvənɛ]. If we choose not to apply (3a), however, rule (3b) will then apply (obligatorily, as we noted), yielding [tüdəvnɛ]. There is thus no way we would expect to obtain the incorrect form *[tüdvnɛ].

What of cases where several schwas are all affected by the same subrule of (2), however? Numerous such examples arise from sequences

of clitics, complementizers, and such items, which in French frequently have the form *Cə* (e.g., *me, te, se, ce, que, le,* etc.). When several of these items are concatenated, interesting results are seen. Consider, for example, the pronunciation of the sequence '. . . *voudrais que ce_que le bedeau* . . .'. We can give the underlying form of this sequence as (4), in which the vowels which could potentially be affected by (3a) are numbered:

(4) /· · · vudrɛ #kə #sə# kə #lə #bədo · · · /
1 2 3 4 5

A form such as this has a large number of possible alternative pronunciations, owing to the fact that the rule's application to any given vowel is optional. We summarize these in the following table, where deletion is symbolized by 'Ø', and nondeletion by 'ə':

(5)

vowel:	1	2	3	4	5
	Ø	ə	ə	ə	ə
	ə	Ø	ə	ə	ə
	ə	ə	Ø	ə	ə
	ə	ə	ə	Ø	ə
	ə	ə	ə	ə	Ø
	Ø	ə	Ø	ə	ə
	Ø	ə	ə	Ø	ə
	Ø	ə	ə	ə	Ø
	ə	Ø	ə	Ø	ə
	ə	Ø	ə	ə	Ø
	ə	ə	Ø	ə	Ø
	Ø	ə	Ø	ə	Ø
	ə	ə	ə	ə	ə

There are a number of possibilities for deleting no vowels, one vowel, two vowels, or three, depending on such factors as emphasis, speed of speech, etc. No possibilities other than those in the tabular example (5), however, are grammatical. The generalization underlying (5) is clear: the rule can apply freely, as long as it does not apply in two adjacent syllables. All of the excluded possibilities involve the deletion of two consecutive schwas, in violation of the well-known 'three-consonant law' of French. Some of the possibilities allowed by (5) would certainly be most unusual, but all can be characterized as potentially grammatical.

It is clear that the simultaneous application algorithm is not capable,

by itself, of dealing with this set of facts. While the optionality of the rule will allow all of the outputs in (5) to be derived, it will also allow the derivation of all of the incorrect possibilities. Dell shows that it is not possible to explain this fact in terms simply of a constraint on the consonant clusters that can arise in French; a cluster such as / . . . kskl . . . / is well formed, as shown by words such as *exclusif,* but the schwa elision rule must still be prevented from operating on (4) to give *[· · · vudrɛkskləb(ə)do · · ·], etc. Some principle must be found either to supplement or to replace the simultaneous application algorithm. It is always possible that there is simply a complex, language particular restriction involved in such a case. It would be much better, however, if we could make this behaviour follow from the general principles of rule application.

One principle that has been suggested which is capable of dealing with the facts of schwa elision in French is that of LINEAR RULES, proposed by Johnson (1970). Johnson proposed that, instead of applying simultaneously, a rule should be applied in one place at a time, proceeding either from right to left or from left to right. The direction to be followed in any particular case is treated as an additional property of the rule, to be specified in each case.[1] In the case of French schwa, we can give rule (2) as above, and specify it as applying from left to right. In this framework, the fact that the rule never applies to two consecutive syllables follows naturally. When the rule applies to any given syllable, it will create a consonant cluster; this consonant cluster will then precede the following syllable, so that when the rule comes to this syllable, the environment of the rule is no longer satisfied, and it cannot apply there. The optionality of the rule at any step will allow it to produce all of the outputs in (5), whereas its left-to-right stepwise character will prevent it from producing any of the incorrect outputs.

The reasoning behind Johnson's suggestion of linear rules is not particularly persuasive in itself. He argues on formal grounds, by analogy to the arguments often given in syntactic studies, that a grammar incorporating simultaneously applying rules has a greater strong generative

[1] A further discussion of this principle of application is found in Howard (1972), where some of the problems of formalism are dealt with. Howard observes that the theory of linear (or 'directional') rules would be more attractive if it were possible to predict the direction of application of a rule from its formal structure, thus removing the necessity of specifying it ad hoc for every rule. As Howard observes, the necessary direction is generally quite clear. It seems unlikely that such attempts to predict directionality can be successful, however, in light of the existence in different languages of formally identical rules which must, in these terms, be applied in different directions. The Slovak rhythmic law and the Gidabal vowel shortening rule to be discussed below constitute such a case.

capacity (in the sense of Chomsky's studies of mathematical models of syntactic structure) than a grammar incorporating only linear rules. That is, he constructs a formal grammar using simultaneous application which generates a language which cannot be generated by any grammar including only linear rules. Whatever the merits of this line of reasoning in syntax, it is not immediately applicable to phonology. Any adequate theory of phonology will have to put very strong substantive constraints on the notion of 'possible rule' (as is suggested in Chapter 15), and these constraints will surely be much more restrictive than any formal constraint of the sort discussed by Johnson. Since the grammars constructed in Johnson's proof contain rules which seem highly implausible as part of the phonology of a natural language, they would undoubtedly be ruled out by substantive considerations, and not on formal grounds. In any case, the correct procedure in this area is surely not to establish such limitations a priori; rather, the structure of individual rules must be established from simple, noninteractive cases, and then the operation of the rule in more complex circumstances investigated. Only on such a basis can we decide what sort of principle should be followed in applying a rule. Johnson's argument from generative capacity for linear rules, then, must be rejected. In the case of French schwa elision, however, we have an argument of more than purely formal character which supports a revision of the theory of rule application such as that envisioned by Johnson. Although any other theory which also accounted for these facts in a natural way would be equally confirmed by them, the theory of linear rules is a perfectly plausible alternative, and clearly superior to the theory of simultaneous application.

We can now enquire as to whether it is possible to construct any other theory which is like the theory of linear rules in being able to deal with French schwa elision, and which is empirically distinct from that of Johnson. If so, we would then look for cases which would allow us to choose between the two. It is reasonably clear what gives the French rule of schwa elision its special character: the problem it raises is due to the fact that no application can be permitted to destroy the environment which was employed for another, simultaneous, application. The various applications of the rule which take place simultaneously, that is, must be independent of one another. This principle recalls the restriction we proposed above in Chapter 12 for Chafe's theory of persistent rules: recall that in that theory all of the rules at a given level were intended to apply simultaneously, and we suggested that it would be plausible to impose as a constraint on the construction of grammars that all members of a set of rules which apply simultaneously in this way must be independent of one another. We might suggest a modification of Chomsky and

Halle's simultaneous application principle, then, which is also subject to this restriction on the simultaneous application of interacting rules. We state such a principle semiformally in the following revision of Chomsky and Halle's principle quoted above:

(6) *To apply a rule, the entire string is first scanned for segments which satisfy the environmental constraints of the rule. As many such segments are identified as can undergo the rule independently of each other, where independence consists in the fact that no segment requires, in the environment which allows it to undergo the rule, the presence of a segment which is itself to be altered by undergoing the rule. After all such segments have been identified in the string, the changes required by the rule are applied simultaneously.*

The notion of independence in this formulation should perhaps be clarified in the following way: if the structural change of the rule is such that it could not introduce a feature specification into an environment segment that contradicts the requirements of the structural description, overlapping applications could be considered independent even if segments undergoing the rule served simultaneously as environments. Suppose, for example, we had a rule which assigns high pitch to a vowel which is neither initial nor final in the word, i.e., in the environment VC_0 ___ C_0V. The structural description of the rule requires only that vowels be present on both sides, and not that they be low pitched; while the rule introduces only the feature of high pitch, and does not alter the status of a segment as a vowel. The modification proposed would allow all of the possible applications of the rule to a string such as $C_0VC_0VC_0VC_0VC_0$ (i.e., the applications of the rule to the second, third, and fourth vowels of the form) to be considered independent. Without this change, however, the application of the rule to the third vowel would be dependent on the applications to both the second and the fourth vowels. We have no empirical evidence to confirm or deny this conjecture, but it seems a plausible refinement of (6) as stated.

Further substance can be given to principle (6) in the form of an algorithm for application of a rule to a given string:

(7) *To apply a rule to a string*

a. Identify all of the segments which satisfy the conditions for the application of the rule, and circle them. For each such segment, identify the minimal substring of

the form which provides the environment allowing the segment to undergo the rule. Each such environment is associated with a particular segment which is to undergo the rule.

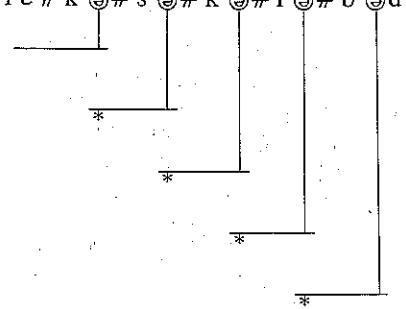
b. *If the rule is optional,*[2] *eliminate from consideration any or all of the segments which could undergo the rule, together with their associated environments.*

c. *If, in the resultant string, any environment contains a circled segment other than the one with which that environment is associated, eliminate the minimal number of segments and associated environments from consideration so as to remove all such violations* [*of principle* (6)].

d. *Apply the change of the rule simultaneously to all of the remaining circled segments.*

Observe the way in which the rule applies to the string in (4). Step (7a) results in the following partitioning of the string, with potential applications encircled and associated environments underlined. Violations of principle (6) are indicated with '*':

(8) /· · · v u d r ɛ # k ◎ # s ◎ # k ◎ # l ◎ # b ◎ d o · · · /

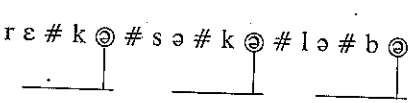


By (7b), we could optionally eliminate any of these circles and their associated underlines. In order to determine the adequacy of the algorithm, however, let us attempt to apply the rule everywhere. At this

[2] Howard (1972) has suggested examples which appear to show that rules may differ as to whether they are optional separately with respect to each potential position of application, or only optional globally (i.e., with respect to application to the whole string or to none of it). Much more study needs to be devoted to the subject of rule optionality, but this topic seems orthogonal to our present concern.

point, when (7c) applies, there are four violations of principle (6). The minimal number of applications which can be disregarded in order to remove these violations is exactly two: the application to the second schwa, and to the fourth schwa. The reader can confirm for himself that if any other application is disregarded instead of one of these two, at least three will have to be disregarded in total in order to eliminate all of the violations of (6). The resultant string, with no remaining violations, is

(9) /· · · v u d r ɛ # k ◎ # s ə # k ◎ # l ə # b ◎ d o · · · /



By (7d), then, the three schwas which remain with circles are deleted simultaneously, giving rise to the output [· · · vudrɛksəkləbdo · · ·], a possible pronunciation of the phrase in question. The remaining possibilities can all be generated by various eliminations of schwas [on grounds of optionality, by (7b)], together with the application of (7c) to force the elimination of others. The reader can readily confirm that all of the possible forms, and none of the impossible ones, can be derived in this way.

After rule (2) has been applied to the string in this way, it cannot of course be allowed to reapply to its own output, since that might permit the deletion of undeletable schwas. Suppose for example that on a first application, we decided to eliminate the first and second schwa from consideration, on grounds of optionality. Two violations of principle (6) would remain, both of which could be removed by eliminating the fourth schwa from consideration. The rule would then apply to delete only the third and fifth schwas, yielding /· · · vudrɛ#kə#sə#k#lə#bdo · · ·/. If the rule were then allowed to reapply, there would be nothing to stop it from applying to delete the second schwa, giving the impermissible /· · · vudrɛ#kə#s#k#lə#bdo · · ·/. It is clear that some rules must be prevented from reapplication to their own outputs, however, so that this property of the schwa-elision rule in French is not a novelty.

It is also clear that cases may arise in which the set of segments which should be eliminated from consideration at stage (7c) will not be uniquely determined by the requirement that the minimal number be disregarded that eliminates all of the violations of (6). Suppose, for example, that we chose to eliminate the first schwa from consideration in (8) on optionality grounds. The three violations of (6) that would still remain could then be removed in two equally minimal ways: by disregarding the second and fourth schwas, applying the rule only to the third and fifth ones; or by disregarding the third and fifth schwas, apply-

ing the rule only to the second and fourth ones. In this case, both outputs are possible, and we might take the theory to predict free variation in such circumstances.

In the case of other, nonoptional, rules we will see that such free variation is not generally allowed. We suggest, however, that choices can be made in most (and perhaps all) such cases by means of the principles we have established already in Chapters 10–12 for the interaction of pairs of rules. Recall that these included the principle of self-preservation; the principle of maximizing feeding (and minimizing bleeding) orders; and the principle of transparency. The three can be arranged in a hierarchy, ordered in just this way; so that each one applies just in case no higher one does. The principle of self-preservation does not, so far as we know, play any role in connection with the application of a single rule [and so in connection with resolving indeterminacies at the stage of (7c)], but the other two do. We will see in examples below that, when a rule is allowed to apply to its own output, an indeterminacy is resolved in favor of the way that will allow one application to feed another rather than to bleed it; while in the case of rules that are not allowed to apply to their own output, and hence could not feed themselves, the indeterminacy is resolved in the direction of maximizing transparency. This use of the same principles of natural order which determine the interaction of a pair of rules to determine the method of application of a single rule is an important result. It allows the principles of natural ordering (and the associated theory of local ordering of pairs of rules) and the revised simultaneous application convention to provide support for one another.

Thus far, we have simply attempted to expound the revised simultaneous application convention, and to demonstrate that it provides a viable alternative to the theory of directional rules in that it can deal with a type of example which the original simultaneous application convention was unable to account for. We have not yet tried to show that there are empirical differences between the revised simultaneous theory and the directional theory. An example which appears to illustrate such differences, and to provide evidence in favor of the revised simultaneous theory, is found in the complex accentual system of Acoma (see Miller, 1965). In this language there are three accents which can appear on a given vowel: a high pitch (marked with an acute accent); a falling pitch (marked with a circumflex); and a 'glottal' falling pitch contour (marked with a glottal stop). The rules for the distribution of these accents are extremely complex, and it is often difficult to determine where accents should be expected to fall, but in a large group of forms, we can be sure of a particular basic accentual distribution. This is the set of forms which contain suffixes conditioning what Miller calls 'accent ablaut'. A group

of approximately twelve suffixes[3] induce this change, which consists in the assignment of the high accent to every syllable of the word (together, in some cases, with the lengthening of the final vowel), except for certain final syllables. Examples of forms with accent ablaut are given in (10)[4]:

(10)
a. *ʔúyáasbáaʔi* 'grinding stone'; cf. *ʔúyâaspaaʔi* 'to grind'
b. *káizúwíita* 'taxpayer'; cf. *kâizúwiita* 'he is paying'
c. *sdyáríiná* 'bells worn by dancers'; cf. *sdári-* 'dangling, jingling'
d. *táaʔawá* 'five times'; cf. *tâaʔa* 'five'
e. *ʔáidyáanáwá gáaná* 'by sevens'; cf. *ʔâidyaana* 'seven'
f. *háanámí* 'the east'; cf. *hâa* 'east'
g. *rúuníšíizé* 'on Monday'; cf. *rûuniši* 'Monday'
h. *héemíšíici* 'Jemez Pueblo'; cf. *hêemiši* 'Jemez Indian'
i. *rúuníšíiʔa* 'every Monday'; cf. *rûuniši* 'Monday'
j. *šíizáačúwání* 'when I woke him up' (suffix *í* + Ablaut)
k. *súwágə́ní* 'when I got dressed'; *suwagə́ni* 'I got dressed'

There are two principle circumstances in which high accents assigned by the accent ablaut rule are lost: one is in short syllables adjacent to a glottalized sonorant, as discussed in Chapter 8 [see also (10d)]. The other condition under which such accents are lost is in a short syllable

[3] It is very difficult to determine what, if any, generalizations can be made about the properties of suffixes causing accent ablaut. While Miller's description is a model of completeness, he frequently goes out of his way to avoid making generalizations. For instance, in discussing the fact that stem final consonants in verbs are preserved before vowels but lost finally and before consonants, he makes it clear that the stem-final consonant will be preserved before any suffix beginning with a vowel. Nonetheless, in discussing the individual consonants which can appear, he makes a point of saying for each that it is preserved before each of a list of suffixes, where this list contains all and only the vowel initial suffixes with which the roots in question are attested. It is only in a footnote that he suggests that there is a principle to these lists (the statement of which would make them unnecessary). He is also quite willing to make use of totally abstract 'process morphophonemes' as the conditioning factor for alternations whose conditioning is not phonetically transparent. It is thus extremely difficult to see how far one could go in explaining accent ablaut in terms of the underlying phonological shape of suffixes which condition it.

[4] Note that the accent on a long vowel or vowel sequence is marked only on the first element.

between obstruents, when followed by another accented syllable. The following examples illustrate this process; each is from a category in which accent ablaut is regular. The vowels whose accent is lost are capitalized:

(11)
a. *kʊbʹóní* 'at sunset'
b. *šɪsíusdyání* 'when I roped him'
c. *ʔúubək'áak'áci* 'nail'
d. *síukAčání* 'when I saw him'
e. *séinúust'Uzími* 'when I put the fire out'

We could write this rule as (12) below:

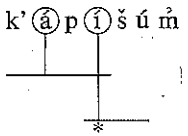
(12) $V \longrightarrow [-\text{accent}] / [+\text{obst}] \_\_ [+\text{obst}]\ C_0 \begin{bmatrix} +\text{syll} \\ +\text{accent} \end{bmatrix}$

Now if two consecutive vowels meet the conditions for the application of this rule, both may be deaccented, as in the forms in (13).

(13)
a. *k'APIšúm̓a* 'every night'; cf. (10i) and *k'ápišu* 'it is night'
b. *ʔAPəgám̓a* 'every evening'; cf. *ʔápəka* 'it is evening'
c. *ʔɪpəkáawání* 'when I chopped wood'
d. *k'Acəkáṅi* 'his cigarettes'

How are we to apply (12) to the forms in (13) in order to obtain this result? Let us see how a form such as /k'ápíšúm̓a/ would be analyzed by step (7a) of the revised simultaneous application procedure:

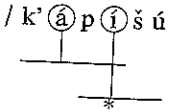
(14)



In (14), there is one violation of principle (6) above. Since the rule (12) is not optional, step (7b) is not relevant, and (7c) must resolve this violation. This can be done by removing either of the two potential applications of the rule from consideration. If we assume that rule (12) is allowed to affect its own output, however, this ambiguity is resolved: if we ignore the second of the two possible applications in (14), the rule will not bleed itself, in that another application will still be possible in the resultant form; while if we ignored the first of the two, the rule would bleed itself, and another application would not be possible. This is illustrated below in two alternate derivations, showing the way rule (12)

analyzes the form at each step:

(15)
a. (7a) / k' á p í š ú m̓ a /

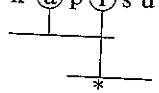


(7c) / k' á p í š ú m̓ a /

(12) / k' a p í š ú m̓ a /
reapplication:
(7a) / k' a p í š ú m̓ a /

(7c) no violations; inapplicable
(12) [k' a p i š ú m̓ a]

b. (7a) / k' á p í š ú m̓ a /



(7c) / k' á p í š ú m̓ a /

(12) / k' á p i š ú m̓ a /
(no reapplication possible)
*[k' á p i š ú m̓ a]

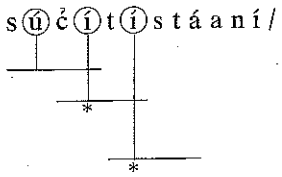
Derivation (15a), in which (12) is allowed to reapply, obtains because the principles of natural order require a bleeding order to be avoided. As mentioned above, the theory of (revised) simultaneous application assumes that this principle applies as well to the resolution of indeterminacies in the application [by (7)] of a single rule as it does to the determination of the relative order of two rules.

The principle under consideration, then, is able to accommodate the examples in (13), on the assumption that (12) is an iterative rule. The theory of linear rules is also able to accommodate these examples, on the assumption that (12) applies from left to right. That is, given /k'ápíšúm̓a/, the first (leftmost) application of the rule yields /k'apíšúm̓a/; then the second (next to the right) application yields [k'apišúm̓a], correctly. We therefore have no way, as yet, of distinguishing between the two theories. It is only when we consider the more complex forms, in which three consecutive vowels meet the conditions for the application of rule (12), that a difference emerges:

(16)
a. *kəẓáçəkáni* 'your cigarettes'
b. *kagácədíní* 'when it was in bloom'
c. *ˌsučítistáani* 'when I was thinking'

Here we see that rule (12) has deleted the first and the third of the three accents which satisfy its structural description, but not the second. Now observe that this is exactly what is predicted by (7). That is, given an input form like /súčítístáaní/, rule (12) will analyze this as (17), at the stage (7a):

(17)



There are two violations for (7c) to eliminate. Both of these can be eliminated if the second possible application of (12) is dropped from consideration, while if either the first or the third is dropped, one of the remaining two will have to be dropped as well. Since (7c) requires that the minimal number of applications be removed from consideration to eliminate the violations of (6), it is clear that only the second possible application should be dropped. However, rule (12) will then affect the first and third vowels, to the exclusion of the second vowel. This gives the correct result, [sučítistáaní], to which the rule cannot reapply. The revised simultaneous application procedure, then, is capable of dealing with forms such as those in (16) satisfactorily.

The theory of linear rules is also capable of dealing with the forms in (16), by assuming that rule (12) applies from right to left. In the input form /súčítístáaní/, the rightmost application of (12) is to the third vowel in the form, giving /súčítistáaní/; in this form, the second vowel no longer satisfies the structural description of (12), and the next possible application to the left is to the first vowel, giving the correct output [sučítistáaní]. Thus, the correct output can be generated by means of linear rules. The difficulty is that the rule (12) must apply from left to right in order to generate the forms in (13), but from right to left in order to generate the forms in (16). This result is inconsistent, and no obvious reformulation of the rules allows it to be avoided. We must conclude, therefore, that the theory of linear rules does not allow for a correct and coherent description of the facts of Acoma accent loss, while the revised simultaneous application theory does. Since both theories account equally well for cases such as schwa elision in French, and there is no other substantive argument in favor of the theory of linear rules, it is probable that we should reject the theory of linear rules in favor of some such device as principle (7).

The revised simultaneous application principle is capable of accounting for patterns such as French and Acoma, and also (as will be discussed below) for simple alternating patterns, in cases where a number of potential applications of a rule overlap in the way we have been considering. One pattern that could not be produced by the algorithm in (7) without additional principles, however, is exactly the pattern that the original simultaneous algorithm was limited to: application of the rule to every possible segment. It is therefore important to inquire as to whether such cases exist in natural languages. There are four cases presently known to us of rules which might appear to present this character. Further examination reveals in each case that an alternative account is at least as adequate, and in some cases to be preferred, in each example. While such primarily destructive discussion is perhaps a bit tedious, each example raises some other issues of interest.

One such example is the behavior of the third tone in Mandarin Chinese, discussed recently by Cheng (1968) and Browne (1972). The surface tones of Mandarin, in isolation forms, are approximately as follows, where contours are given in Chao's tone letters[5]: Tone 1 is phonetically high level ([˥]); tone 2 is mid–rising ([˧˥]); tone 3 is falling–rising ([˨˩˦]); and tone 4 is high–falling ([˥˩]). The interesting problems arise in connection with the behavior of sequences of tone-bearing syllables, where rules of tone sandhi apply. Sandhi affecting tone 3 is of particular interest: when followed by another instance of tone 3, tone 3 is replaced by tone 2; and when followed by another tone other than tone 3, tone 3 is replaced by a mid-falling contour, the so-called 'half–third' tone ([˨˧]).

The important fact to observe is the behavior of several consecutive instances of tone 3. Cheng (1968) discusses the influence of syntactic structure on the realization of such sequences; we can assume that processes determined by the interaction of syntactic structure with speed of speech and other stylistic concerns delimit stretches of the utterances within which sandhi can occur, and at the ends of which the tones display their isolation, or word-final values. Within such a sandhi-domain, all consecutive instances of tone 3 except the last are converted into tone 2, as required by the above sandhi rule. Thus, a sequence *3–3–3–3–3#* becomes *2–2–2–2–3#*. The obvious suggestion is that the sandhi rule is something like the following:

(18) [tone 3] ⟶ [tone 2] / ___ [tone 3]

[5] There is some question as to the exact phonetic values in some cases; it is unclear, for example, whether the final high pitch in the isolation form of tone 3 is to be identified with the level of tone 1 and/or the final high of the isolation form of tone 2, or is to be marked lower. These details are irrelevant to our analysis, and we will assume that the contours can be adequately specified in terms of high, mid, and low pitch.

If (18) is, in fact, the correct form of the rule, it is necessary to apply it in a way which violates principle (6) above, since some of the instances of tone 3 before which sandhi of other tone 3 syllables is to apply are themselves being converted into tone 2 by the same rule at the same point in the derivation. Since each instance of tone 3 is altered to tone 2 under the influence of a following instance of *tone 3*, which may itself be changed, it is difficult to see how the independence of all of the consecutive applications can be maintained. Notice that the simultaneous application principle of Chomsky and Halle can handle this situation perfectly well (indeed, this is the only possible behavior, on the basis of that algorithm); the linear rules theory can also handle this rule, by declaring it left to right.

Upon further analysis of the tone system, however, this problem disappears. As will be discussed below in Chapter 14, several writers have recently argued that a single syllable can contain several separate tone elements, realized in sequence, and that the so-called contour tones are not elementary units, but rather sequences of level tones. The Mandarin 'mid–rising' tone, for example, is not to be treated as an unanalyzable unit, but rather as a sequence '*m–h*' (*l* = low, *m* = mid, and *h* = high). Similar treatment should also be given to the other contour tones. Let us assume now that the tone contours for Mandarin syllables are specified in terms of two tone levels per syllable, in sequence: tone 2, mid-rising, is as just noted represented as [*m–h*]; tone 4, high-falling, is represented as [*h–l*]; and tone 1 (high level) is agreed to be simply [*h–h*]. In the case of tone 3, however, several investigators have argued (completely independently of the issues we are concerned with here) that this tone has an underlying representation which is distinct from any of its surface shapes. Both Cheng (1968) and Woo (1969) have argued on other grounds that tone 3 should be treated as underlyingly a low level tone ([˩] or [*l–l*]). In that case, we would need rules quite distinct from (18) to predict the phonetic forms. Let us assume now that (at least in tone representations) syllable boundaries are marked, so that a rule can determine whether two adjacent tone elements belong to the same syllable or to different syllables. We will mark syllable boundaries in the derivations below with a period (.).

In tone 3 syllables, we can note that the first tone element is realized not as the underlying low, but as mid in all cases. To account for this, we need a rule such as (19):

(19) $$[l] \longrightarrow [m] \; / \; . \_\_$$

This formulation is adequate, since tone 3 is the only tone in the system which begins with *l* in the underlying form. Taken by itself, rule (19) will

correctly generate the 'half–third' contour ([˨˩]), though if no other rules appeared, it would incorrectly assign this value to all instances of tone 3 in all positions. We need, in addition, principles by which the other contours can be produced.

The isolation value of the tone is easy to deal with. What is necessary is to assume (with other recent authors, as will be discussed in Chapter 14) that some syllables in a language may come to have a different number of sequentially realized tone elements assigned to them than others. In that case, we can say that the isolation (or final) form of tone 3, mid–falling–rising ([˧˩˧]), is to be represented as [*m–l–h*], and that syllables bearing this tone contain three tone elements, rather than two. The rule which produces this contour, then, is simply a process which inserts an extra tone element ([*h*]) into a tone 3 syllable followed by the boundary of the phrase (i.e., when not followed in the phrase by another tone):

(20) $$\emptyset \longrightarrow [h] \; / \; .[-h][l] \_\_ .\#\#$$

The term '[−*h*]' in this rule means either mid or low, and is a perfectly natural one in a system of features such as that of Halle and Stevens (1971), where it would be expressed more accurately as '[−stiff vocal cords]'.

We have ignored here the problem of a small class of fixed compounds with an initial syllable of tone 3 and a second syllable having the 'neutral' tone. Tone 3 is always said to be realized in its isolation form ([˧˩˧]) in these compounds. It is interesting to note, however, that where the complex (or three element) realization of the third tone is followed by a neutral tone syllable, the neutral tone is realized as high, and the tone 3 syllable is in fact realized as *m–l*. What is clearly going on in this case is that the contour is spread out over more than one syllable if it contains more tone elements than usually can fall on one syllable (in Mandarin, two) and an extra syllable is available to accomodate the surplus. This is exactly analogous to phenomena in African and other tone systems which are discussed in Chapter 14.

Rules (19) and (20), then, will correctly derive the third and half-third contours, which phonetically realize the underlying low-level tone. But what of our original problem, the generation of a second tone from a third followed by another third? For this, we need simply specify a rule which converts the second *l* of a syllable to *h* when it is followed by another syllable-initial *l*:

(21) $$[l] \longrightarrow [h] \; / \; .[l] \_\_ .[l]$$

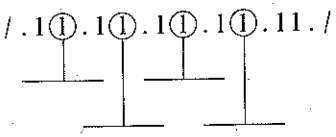
Rule (21) applies before rule (19) converts the first element of a low level tone to [*m*]; this ordering follows from the fact that (19) would otherwise bleed (21). Now given a sequence of several third tone syllables, we propose the following derivation:

(22)

underlying form:	/ . l l . l l . l l . l l . l l . # /
rule (21):	/ . l h . l h . l h . l h . l l . # /
rule (19):	/ . m h . m h . m h . m h . m l . # /
rule (20):	/ . m h . m h . m h . m h . m l h # /
phonetic output:	[˥ ˥ ˥ ˥ ˧ #]

Now observe that in connection with this system of rules, the problem of the application of the third-tone sandhi rule has disappeared. It is true that several instances of third tone are converted into second tone, but the crucial difference between rule (18) and rule (21) is the following: in (21), the element which undergoes the rule is a syllable-final tone element; the elements which furnish the environment for the operation of the rule are all syllable-initial elements. Thus, the rule is never required to make use of an item in satisfying its structural description which is to be affected by the structural change of the rule. In terms of the algorithm (7) above, rule (21) analyzes the underlying form in (22) as follows:

(23)



Since no circled element appears in an underlined environment which is associated with any other circled element, the rule can (and must) be applied to all of the circled elements simultaneously. This example is thus seen not to be inconsistent with principle (6) and its realization as the algorithm (7), as might be suggested by a rule such as (18) which is not based on a complete analysis of the internal structure of tones.

Another example which appears to contradict principle (6) is discussed by Browne (1970, 1972) in Slovak. In Browne (1970), it is demonstrated that Slovak is in general subject to a 'rhythmic law', by which a long vowel is shortened following another long vowel:

(24) $$[+\text{syll}] \longrightarrow [-\text{long}] \,/\, \begin{bmatrix} +\text{syll} \\ +\text{long} \end{bmatrix} C_0 \,\_\_$$

This rule applies to, e.g., underlying /čīt+ām/ to give [čītam] 'I read (pres.)' (for the length of /+ām/, cf. [volām] 'I call', with a short root vowel). In general, only two long vowels in succession can arise in Slovak, but there is an exception in forms which contain the iterative marker /+āv+/. This can appear, for example, in the first singular from the root /čīt/, giving underlying /čīt+āv+ām/, with three consecutive long vowels. Rule (24) applies to shorten both of the vowels which meet its environment, giving [čītavam] 'I read (pres. iterative)'. Thus far, this case is parallel to that of the accent-loss rule in Acoma, and we can conclude that rule (24) is allowed to apply to its own output. The simultaneous theory of Chomsky and Halle (1968) can also handle these data, as can the theory of linear rules [which would characterize (24) as applying from right to left].

The problem arises from the fact that there is one possible source for a sequence of four or more consecutive long vowel syllables in Slovak: the iterative marker can itself appear several times, indicating an increased notion of iteration in forms such as /čīt+āv+āv+āv+ām/ 'I read and read and read and read (pres.). The difficulty for principle (6) is that all of the long vowels (except, of course, the first) in this form are shortened, giving the output [čītavavavam]. This result is consistent with either the simultaneous or linear theories, but not, apparently, with principle (6). The algorithm (7) should give something like [čītavāvavam], with an alternating pattern of length.

As long as forms with multiple instances of the iterative marker are the only instances of four or more long vowels, however, the Slovak rhythmic law cannot be considered a critical case. As is evident from the gloss of [čītavavavam], 'I read and read and read and read (pres.)', English, too, has a device for forming intensified iteratives: copy the verb several times. This device has some very interesting properties, as has been pointed out by Browne (in personal communications and class discussions in 1967). There are some verbs in English for which two past tense forms exist in free variation for many speakers: the past of *dive* can be either *dived* or *dove*, the past of *leap* can be either *leaped* or *leapt* (i.e., with or without shortening of the stem vowel), etc. Any of the forms in free variation can be the basis of an iterated formation, so we can have either 'he leapt and leapt and leapt' or 'he leaped and leaped and leaped'; but the two cannot be intermixed, so we cannot have *'he leapt and leaped and leapt', etc. This suggests that the formation of iteratives does not take place freely within the syntax, but rather takes place only after inflectional elements have received their phonological realization, in order to insure that all of the iterated elements will have the same phonetic shape. If we extend this constraint analogously to the case of Slovak repeated iteratives, we might suggest that parallel facts obtain there. Suppose that /čīt+āv+āv+āv+ām/ is not the underlying

form of [čītavavavam], but rather /čīt+āv+ām/ is. Rule (24) applies (and reapplies) to this representation, giving /čīt+av+am/. If the repetition of the iterative element is carried out only at this point, then the element which is copied will be not /+āv+/, but /+av+/. Thus, if an intensified iterative is formed by copying an element whose phonetic shape has already been determined, as is suggested by the formation in English, forms such as *čītavavavam* can be derived without any violation of principle (6).

Another case that appears to contradict principle (6), brought to our attention by Paul Kiparsky, is the rule of consonant gradation in Finnish. It will be remembered from Chapter 6 above that this rule affects stops at the beginning of a short, closed syllable, simplifying geminates and converting single stops into 'weaker' segments. Since the presence of a geminate closes the preceding syllable, this raises the possibility that an element which is providing the environment for one application of the rule may itself be subject to another application. Recall that the environment for the gradation rule was stated as[6]

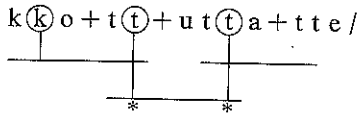
(25) $$\begin{bmatrix}-\text{cont}\\ \alpha\text{cor}\end{bmatrix}\left[\frac{\quad}{\alpha\text{cor}}\right][+\text{syll}][-\text{syll}]\left\{\begin{matrix}[-\text{syll}]\\ \#\end{matrix}\right\}$$

The last element of a geminate, then, might be analyzable both as the affected segment in one application of the rule, and as the final segment of the environment in some other application. Owing to the extensive possibilities of suffixation in Finnish, such cases are not too hard to come by. From the word *rokko* 'infectious disease, esp. one producing skin pustules', we can form a verb *rokottaa* 'to inoculate' by adding the suffixes /+tta+/ 'to fix up with something' (cf. *vero* 'tax', *verottaa* 'to tax') and /+ta/ 'infinitive marker' (which loses its *t* after a short vowel). The first person singular of this verb is *rokotan,* showing gradation both of the root and of the suffix. From *rokottaa* we can further form a causative *rokotuttaa* 'to cause to be inoculated' with the suffix /+utta+/ (cf. *kirjoittaa* 'to write' versus *kirjoituttaa* 'to cause to be written'; /+utta+/ causes the loss of an immediately preceding vowel by regular processes). The underlying structure is /rokko+tta+utta+ta/. The second person plural of this verb then shows four consecutive geminates, each separated by a short vowel: /rokko+tta+utta+tte/ gives *rokotutatte*, with the first three of these geminates undergoing gradation.

[6] We confine our attention here to the portion of gradation which affects geminates. The problem being dealt with could not possibly arise in the case of gradation of single stops.

Forms like *rokotuttaa,* with two overlapping instances of gradation, can be reconciled with any of the theories we have considered. The theory of simultaneous application obviously needs no modification for this example; the linear rule theory deals with it by characterizing gradation as left-to-right; and the revised simultaneous theory incorporating principle (7) describes the rule as one which reapplies to its own output.[7] A case like *rokotutatte,* however, is different. The simultaneous and linear theories need no modification, but the example is inconsistent with principle (6) if (25) is the correct formulation of the environment. Algorithm (7), that is, would analyze this form in terms of the gradation rule (which follows the loss of *a* before *u*) as (26):

(26)



There are, as we see, two violations of (6) in (25), which (7c) would eliminate by disregarding the second of the two possible environments for gradation. This would result in the incorrect form **rokottutatte,* with only the first and third geminates reduced.

It is not hard to argue, however, that (25) is incorrect as the statement of the environment for gradation. The rule, as is generally agreed, applies at the beginning of a short closed syllable; this is formulated as (25) simply because there is at present no device available in phonology to permit rules to refer directly to syllable structure. There is an increasing body of evidence, however, that syllables must be delimited in phonological representations. This matter is discussed below; an example of a rule making reference to syllable boundaries is the Mandarin tone process discussed previously, and other such rules are suggested in Chapter 14 and by Hooper (1972). The term '$\{\frac{C}{\#}\}$' is present in rule (25) simply in order to encode the principles of syllabification in Finnish into the gradation rule. This last element of the structural description is necessary not because of any property it has itself, but simply because its

[7] It is of course true that single stops produced by gradation must not be further weakened by the gradation rule. Thus, *rokottaa* must not be further converted to **roottaa*. This is perfectly consistent with the provision that the rule applies to its own output, however; recall that in Chapter 9 we imposed the general limitation that rules never reapply in the same position as a previous application. Furthermore, as we argued in Chapter 6, gradation of single stops is a distinct rule from gradation of geminates, and regardless of the behavior of degemination, the single stop rule precedes the geminate rule (by the principle suggested above in Chapter 11, in connection with the Gitksan rules converting *x* into *h* and *h* into ∅ intervocalically, if that principle can be formally founded).

presence will cause a syllable boundary to fall after, and not before, the consonant preceding it, thus closing the syllable. The rule should really be formulated in such a way as to refer to a short syllable ending in a consonant, and should not have to duplicate the statement of the conditions under which this will be the case:

(27)[8] $\begin{bmatrix}-\text{cont}\\ \alpha\text{cor}\end{bmatrix}\left[\begin{array}{c}\\ \alpha\text{cor}\end{array}\right]$ [+syll][−syll].

If the environment of gradation is given as (27), however, the problem this example would pose for principle (6) disappears. In this case the three separate examples of degemination in *rokotutatte* are independent in the relevant respect. The revised gradation rule now analyzes the form not as (26) but as (28):

(28) /.r o k.ⓚ o + t.ⓣ + u t.ⓣ a + t.t e./

Although the environments overlap in (28), principle (6) is not violated, since no segment which is affected (i.e., no circled segment) is used in an environment other than the one it determines. Accordingly, (7c) is inapplicable, and the rule applies so as to delete all three consonants which are circled. Subsequent resyllabification will relocate the syllable boundaries in front of the remaining members of the original double stops, giving [ro.ko.tu.tat.te] as the output. We see, however, that Finnish consonant gradation is not in fact a counterexample to principle (6).

A fourth and final example, which appears to contradict (6), is provided by the rule(s) of Slavic phonology known as Havlík's law. Common Slavic possessed two lax high vowels Ь and Ъ, known as the

[8] The symbol '.' in this rule represents syllable boundary, as above. Rule (27) is essentially the formulation of the environmental specification for Finnish gradation which is suggested independently by Hooper (1972). Both Hooper's discussion and the present one ignore the fact that problems are presented for any analysis of gradation by postvocalic [j], since some occurrences of this segment close a syllable, others do not, and others (while they do not close the syllable) do not make the syllable long, either. These problems, which have not been completely resolved, are probably irrelevant to the present concerns; if anything, they seem to suggest that it may be even more important to recognize the reality of syllables in formulating the rule. It can be noted that Hooper's formulation of the environment without use of syllable boundaries is a bewilderingly complex one, that makes the difference between (25) and (27) look greater than it has to be.

JERS. In all of the modern Slavic languages, however, these vowels have been replaced phonetically. The conditions for this replacement were first formulated (for Czech) by Havlík, in a way that appears to be valid for early West Slavic in general. The jers can be divided into two classes: weak jers are those that are final or followed by a non-jer vowel. Other jers (i.e., those followed by another jer) are strong. Havlík's law states that weak jers were lost, and strong jers lowered to *e* and *o*, depending both on the original backness of the jer in question and on the following consonant.

Isačenko (1970) provides an excellent and comprehensive discussion of the history of these phenomena in Russian. He first disposes of the claim that Ь was uniformly lowered to *e*, and Ъ to *o*, and that *e* was later backed to *o* in forms like[9] /p,os/ (from *pЬsЪ*). He formulates the following as the treatment of the strong jers:

(29) /Ь/ ⟶ /e/ before soft consonants only
/Ь/ ⟶ /o/ in all other cases
/Ъ/ ⟶ /o/

We will refer to the process in (29) as simply 'jer → [−high]' without attempting to express it in formal terms. We will also follow Isačenko in representing both jers indifferently as *, ignoring (incorrectly) the difference between them in general.

From the etymologies of forms that do not show alternations, and from attested forms shortly after the period to which the 'jer-shift' is assigned (late twelfth to early thirteenth century), it appears that Havlík's law was a valid synchronic rule at that time, in essentially the form outlined above. The rule is stated by V. Kiparsky (1963; cited by Isačenko, 1970) as follows: "starting with the end of the word, or with the last full vowel, every odd Ъ, Ь is dropped, while every even Ъ, Ь is vocalized" (or lowered). The rule is illustrated with such forms as

(30) a. /l Ь s t Ь ts Ь / ⟶ /l (Ь) c t e ts / (R) 'flatterer'
1 2 3

b. /r Ъ p Ъ t Ъ / ⟶ /r p o t (Ъ) / (0R) 'murmer, grumble'
1 2 3

c. / s Ъ n Ь m Ъ / ⟶ /s n e m (Ъ) / 'meeting, gathering'
1 2 3

[9] In accordance with Slavistic practice, we use a following comma to indicate the palatalized ('soft') consonants in transcriptions. In addition, the sign Ь indicates the front jer, and Ъ the back jer.

In these cases, odd-numbered jers (from the end of the word) are dropped, though sometimes preserved orthographically to represent the quality of a preceding consonant; and even numbered jers are lowered. This process could be formulated in either of two ways:

(31) a. jer $\longrightarrow \emptyset$ / ___ $C_0 \left\{ \begin{bmatrix} \# \\ +\text{syll} \\ \{-\text{high} \\ +\text{tense}\} \end{bmatrix} \right\}$

b. jer $\longrightarrow$ [−high]

(32) a. jer $\longrightarrow$ [−high] / ___ $C_0 \begin{bmatrix} +\text{syll} \\ +\text{high} \\ -\text{tense} \end{bmatrix}$

b. jer $\longrightarrow \emptyset$

The difference between (32) and (31) rests on the choice of which process to treat as the 'elsewhere' case. If the lowering of jers is treated as the elsewhere case (i.e., as specifying the fate of residual jers after the operation of jer-loss), (31) is the formulation; if the loss of (nonlowered) jers is treated as the elsewhere case, (32) results. The principal difficulty in providing a solution to this example comes from the fact that there is apparently no evidence that points to a choice between (32) and (31). From historical sources, it appears that weak jers were lost in certain positions before other aspects of the jer-shift occured, which would favor (31) slightly, but this evidence is virtually impossible to interpret unambiguously. The derivations in (30), however, would suggest that (assuming no other factor, such as intonation, differentiated between strong and weak jers) (32) is correct, given our assumptions thus far about the application of rules in forms to which they can apply in several places. On the theory of linear rules, (32a) applying from right to left (as suggested by V. Kiparsky) followed by (32b) gives the correct result; while on the revised simultaneous theory, (32a) prevented from reapplying to its own output, followed by (32b), will also generate the required alternating pattern in a fashion analogous to other examples discussed in this chapter. There is no obvious way the original simultaneous theory can provide this result from either system of rules, nor is there any obvious way (31) can be applied correctly on any theory.

In Modern Russian, however, the facts are somewhat different. As first proposed by Lightner (1965a), later discussed by Worth (1967), and supported by Isačenko (1970), the 'Havlík's law' rule now applies quite

differently. Worth's formulation was that the rule should now be treated as applying from left to right, rather than from right to left, as originally. Isačenko phrases this as follows: 'In stems not divided by a prefix / preposition boundary (=) and containing more than one vowel / zero alternation (i.e., more than one underlying jer: *sra*), only the last of these alternations is preserved; all other vowel / zero alternations are vocalized.' The import of this is as follows: any jer followed by another jer within the stem is lowered. Only the stem final jer alternates, being either dropped or lowered depending on the following desinence. This can be illustrated by forms such as the following:

(33) a. from /v,etr/, by epenthesis in stem-final *TR* clusters:

/v,et,∗r/ from which /v,et,∗r+∗k+∗/ gives *v,et,erok* and /v,et,∗r+∗k+a/ gives *v,et,erká* from which further, /v,et,∗r,+∗č+∗k+∗/ gives *v,et,eróček* and /v,et,∗r,+∗č+∗k+a/ gives *v,et,eróčka*

b. from /st∗kl/, by epenthesis, /st∗k∗l/; from this /st,∗k∗l,+∗n+ɨj/ gives *st, ekól, nɨj*

This result could be obtained by applying rules (31) on the simultaneous theory; or by applying (32) from left to right, on the linear rules theory. It could also be obtained by applying rules (31) on the revised simultaneous theory. The change from the original alternating pattern of jer loss to the Modern Russian one of the preservation of all but the last jer would thus be viewed on the linear theory as a change in the direction of the jer-lowering rule; while on the revised simultaneous theory it would be viewed as the change from (32) to (31). Either change could be motivated by the fact that (as Isačenko notes) it results in greater uniformity of the forms within a single inflectional paradigm, since the alternation between vowel and zero is now confined to the last syllable of the stem. These facts, then, do not pose any particular problem for either of the theories under consideration. The reason for citing them is that the rule is often taken in works on Slavic to be (32), and in that case the Modern Russian rule would present a problem for the revised simultaneous theory. There do not appear to be any facts, however, which require (32) as the formulation, and (31) (which is the form assumed by Halle, 1971) is just as adequate. (31), as we saw, removes the problem for the revised simultaneous theory: the rule which must apply in several places on this analysis (31b, jer-lowering) is unconditioned, and hence the question of violations of (6) cannot arise.

Other facts show that this analysis is at least incomplete. The formulation given above covers only cases of stems not containing the boundary =; when forms with this boundary are considered, a different pattern emerges. If a jer precedes this boundary, and is followed by another jer after it, the preservation of the first jer is determined by the fate of the second. Thus, we have alternations such as (34):

(34) a. /s∗=ž∗g+∗/ gives *sžog;* /s∗=ž∗g+la/ gives *sožgla*
b. /ot∗=p,∗r+∗/ gives *otp,er;* /ot∗=p,∗r+ū/ gives *otopru*
c. /v∗=r∗t+∗/ gives *v rot;* /v∗=r∗t+ū/ gives *vo rtu*

The alternating patterns here are like those of the original Havlík's law cases. They could be provided for in several ways, none of which are satisfactory. The simplest, and least insightful way to provide for such cases is simply to introduce another rule which precedes (31) [or (32)], which drops a jer followed immediately by prefix / preposition boundary exactly when the next two syllables contain jers:

(35) $\text{JER} \longrightarrow \emptyset \,/\, \_\_\_ = C_0\text{JER}C_0\text{JER}$

This brute force method will work, but hardly explains anything. Another unpleasantly brutal solution would be to say that, on the linear theory, rule (32a) applies from left to right within a stem, but (simultaneously) from right to left across a = boundary. Yet another alternative which surely must be rejected is to say, on the revised simultaneous theory, that rules (31) apply within the stem, but rules (32) apply across =.

Clearly, if there is a generalization to be captured here it is the following: if the following vowel is or becomes a full vowel, the prefix or preposition vowel is lost, as any jer is lost before a full vowel. If the jer of the stem is not turned into a full vowel, however, the jer in the prefix or preposition is lowered, as any jer is lowered before another jer. This fact cannot be obtained by any simple application of the rules and theories discussed above, and it is necessary either to lose it [by adopting rule (35) or something like it] or to modify our theories somewhat. The most attractive modification would seem to be the following: suppose we adopt a version of the theory of cyclic rules, according to which integral lexical items (i.e., stems plus derivational affixes and inflectional affixes but not clitic elements such as prepositions) undergo a subset of the rules of the phonology as lexical items, in the dictionary as it were. Then the full word, with inflectional and clitic elements attached, undergoes all of the rules of the phonology. This position can perhaps be supported by some other observations of Halle's (1973) about stress in Russian, though these would have to be formalized more completely before their import could be properly assessed. On this analysis, the rule of jer-lowering (32a) would apply both at the lexical item level and at the level of the full phonological word. In that way, the quality of the stem jer could be determined before that of the prefix or preposition, as seems to be indicated. This theory would thus capture the relevant generalization, but only at the cost of a major theoretical innovation which cannot be said to have important support elsewhere. Until such support is forthcoming, the unpleasantly ad hoc rule (35) would seem to be unavoidable. In any case, this problem is unconnected with the other problems discussed above, and whatever solution is adopted for it can be incorporated equally well into the grammar with either linear rules (32) or with rules (31) applying under the revised simultaneous convention.

We have seen, therefore, that none of the examples which have been discussed in the literature are, in fact, counterexamples to principle (6). In each case, an alternative analysis is available, for which some evidence can usually be given, such that principle (6) is satisfied. This absence of counterevidence furnishes definite evidence in favor of principle (6), for it makes it clearer that adjacent applications of a rule are possible exactly when they can be treated as independent, as principle (6) requires. No alternative analysis exists, however, for the stress-loss rule in Acoma in terms of either the simultaneous or the linear application theories. We should, accordingly, prefer the revised simultaneous application principle to any of these.

This theory produces cases such as that of Mandarin, etc., where the rule in fact applies everywhere it could simultaneously, and also cases such as French, where the rule applies only in alternate positions, depending on whether adjacent applications are independent [in terms of (6)] or not. By allowing the rule to apply to its own output, and then invoking the natural ordering principle of minimizing bleeding application, it can also accommodate cases like that of Acoma, where two adjacent applications are possible, but where three possible applications result in an alternating pattern. Other cases also exist, in which three possible overlapping applications give rise to alternating patterns, as in Acoma, but where two overlapping applications also give rise to such a pattern, unlike the Acoma case. It is to such an example that we now turn.

In the Australian language Gidabal,[10] there is a process of vowel

[10] The data below on Gidabal are from Geytenbeek and Geytenbeek (1971), as cited by Kenstowicz and Kisseberth (1972).

shortening which is formally just like that of the Slovak rhythmic law: it applies to long vowels that follow other long vowels:

(36) $$[+\text{syll}] \longrightarrow [-\text{long}] \; / \; \begin{bmatrix} +\text{syll} \\ +\text{long} \end{bmatrix} C_0 \; \_\_$$

The following alternations attest to the operation of this rule in the language:

(37) a. *badi+ya:* 'should hit'; *badi+ye* 'may hit'
b. *yaga:+ya* 'should be fixing'; *yaga:+ye* 'may be fixing'
c. *ga:da+ya:* 'should chase'; *ga:da+ye* 'may chase'
d. *mani+ya:gan* 'to get wallaby'; *muru:n+dagan* 'to get firewood'

Clearly, we have to do here with roots such as /badi/ 'hit', /yaga:/ 'be fixing', /ga:da/ 'chase', /mani/ 'wallaby', and /muru:n/ 'firewood'; and suffixes such as /+ya:/, /+ye/, /+Ca:gan/. Long vowels in the affixes are shortened when possible by rule (36). There are also cases in the language of affixes which cause the lengthening of a preceding vowel (along with the lowering of /i/ to /e/). When the preceding vowel is already long, it stays that way; but when it is preceded by another long vowel, it is not lengthened:

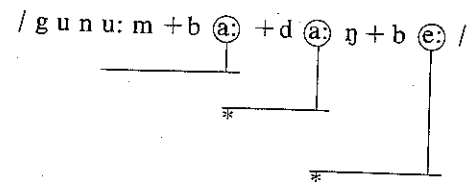
(38) a. *badi+ye* 'may hit'; *bade:+nj* 'will hit'; *bade:+n* 'hit (past)'
b. *yaga:+ye* 'may be fixing'; *yaga:+nj* 'will be fixing'; *yaga:+n* 'was fixing'
c. *ga:da+ye* 'may chase'; *ga:da+nj* 'will chase'; *ga:da+n* 'chased'

The difference between (38c) and the others is also predicted by rule (36).

For cases in which four consecutive long vowels arise, the last three of these can potentially undergo the rule. Consider, for example, the underlying form /gunu:m+ya:+da:ŋ+be:/, consisting of the root /gunu:m/ '(on the) stump', with the suffixes *-ya:* (seen also, with assimilation of the /y/ to preceding nasal consonants, in *bala+ya:* 'is under' *ba:m+ba* 'is halfway'), *-da:ŋ* [also found in *njule+da:ŋ* 'he (emph.)', *yu:+daŋ* 'much later'], and *-be:* (also seen in *gadi+be:* 'right here', *buřu:ř+be* 'is certainly two'). This form gives the surface *gunu:m+ba+da:ŋ+be* 'is certainly right on the stump', with the shortening of the second and fourth, but not

the third, of the long vowels in the underlying form. This is just what we would expect: rule (36) gives an analysis of this form as (39) after step (7a) of the algorithm for applying a rule applies:

(39)



In (39), there are two violations of (6), which are eliminated by (7c) by disregarding the second of the three possible applications. This is the minimal way to eliminate both violations, as was the case in previous examples. The rule then applies to the remaining circled vowels (the third and fifth of the entire form), giving the correct result *gunu:m+ba+da:ŋ+be*. This alternating pattern is consistent with either (7) or with the theory of linear rules (which characterizes (36) as left to right in Gidabal, while the same rule operates right to left in Slovak), but not with the original simultaneous application theory.

Consider what happens when three (and not four) consecutive long vowels occur. This arises when we consider a formation just like that of (39), with a root whose last vowel is basically short: /djalum+ba:+da:ŋ+be:/, which gives rise to *djalum+ba:+daŋ+be:* 'is certainly right on the fish'. Note that here only the middle one of the three vowels that are underlyingly long is shortened. Similarly, the present tense suffix *-la* requires a preceding vowel to be long (as did *-nj* and *-n* considered above), and can be added to bases which have already been suffixed with *-le* 'repetitive'. When these two are added to a stem with final short vowel, the repetitive affix is long and the present affix is short; but if the last vowel of the stem is long, the distribution of length in the affixes is reversed:

(40) a. *nama+le:+la* 'is holding'; *badi+le:+la* 'is hitting'
b. *yaga:+le+la:* 'is fixing'; *ginjalga:+le+la:* 'is coughing'

These facts obviously suggest that *-la* itself has an underlying long vowel, as well as a lengthening effect on the preceding vowel; and that the length in /-la:/ only shows up in cases where the induced length in the preceding vowel is canceled. The underlying form of 'is fixing', then, is /yaga:+le:+la:/ (after operation of the morphological lengthening rule). The vowel shortening rule then has two overlapping possibilities,

but only the first of these is actually applied, just as was the case in *djalum+ba:+daŋ+be:*.

This result is inconsistent with the original simultaneous theory, just as in the case with more than two overlapping applications; it is consistent with the linear theory (since the same left-to-right application of (36) will give the right results in all cases). In terms of the revised simultaneous application convention, a form like /djalum+ba:+da:ŋ+be:/ is analyzed as (41):

(41) / dj a l u m + b a: + d a: ŋ + b e: /

*

There is one violation of (6) in this form at the point (7c) applies, which can be eliminated by disregarding either of the two possible applications. If we assume that the rule applies to its own output, the first will be disregarded, and the rule will apply only to the fifth vowel; it will then reapply, shortening the fourth vowel and yielding the incorrect form **djalum+ba:+daŋ+be*. Let us suppose that the rule is not permitted to apply to its own output: in that case, the principle of maximizing the possibility of reapplication (minimizing bleeding) is irrelevant, and we are left with a choice of ways for eliminating the violation of (6) in (41). In that event, however, the principle of minimizing opacity allows us to make a decision: if we disregard the first of the two possible applications, we will derive *djalum+ba:+da:ŋ+be*, which is opaque in that it contains an instance of a long vowel following another long vowel, which should be eliminated by (36). However, if we disregard the second possible application instead, we derive *djalum+ba:+daŋ+be:*, which is not opaque because the long vowels are not adjacent. The minimizing of opacity, then, requires us to disregard the second possibility, and this results in the correct form. This example too, then, is consistent with the revised simultaneous application convention. Whereas the linear theory characterizes the difference between the vowel shortening rules in Slovak and in Gidabal in terms of the ad hoc parameter of right-to-left versus left-to-right application, the revised simultaneous theory characterizes this as the difference between rules that can and rules that cannot reapply to their own output, a distinction which must be made in any event.

Of the theories that have been considered in this chapter, the simultaneous theory adopted by Chomsky and Halle (1968) is clearly the least adequate, since it cannot deal with several of the cases above. The linear theory is a distinct improvement, since it can deal with all of them except the accent loss in Acoma. The revision of the simultaneous application theory [incorporating (6) and (7)] must, however, be considered the most nearly adequate on the basis of evidence presently available, since it can also accommodate the Acoma example, which provides rather striking confirmation of its claims. It has the important merit of making use of the same principles that govern the interaction of pairs of rules to control the interrelation of applications of the same rule. The desirability of such a solution has been pointed out by Kenstowicz and Kisserberth (1972) (who do not really claim, however, to provide a satisfactory one). The issue cannot be considered completely closed, (if any phonological issue ever can), due to the paucity of critical examples. Although interesting, the Acoma example is isolated; and the absence of cases in which four, five, and more applications of a rule can impinge on one another (and where the rule can reapply to its own output) leaves room for doubt. At present, however, it is safe to say that there is no positive evidence for the theory of linear rules, and a certain amount of evidence against it, favoring a revision of the simultaneous application principle along the lines of (6).