Competing Changes as a Cause of Residue

Phonological change may be implemented in a manner that is phonetically abrupt but lexically gradual. As the change diffuses across the lexicon, it may not reach all the morphemes to which it is applicable. If there is another change competing for part of the lexicon, residue may result. Several fundamental issues in the theory of phonological change are raised and discussed.

In the literature on sound change, much has been made of the neogrammarian doctrine that sound changes operate without exceptions. Without some such hypothesis any description would be a long list of unsystematic correspondences, with no assurance that the same sound under comparable conditions would not change into a variety of different sounds, with no governing principle whatsoever. This point of view, which Hockett has termed the "regularity hypothesis," has been richly rewarding in historical research; as a celebrated part of the heritage of modern linguistics, the point needs no elaboration here.

When irregularities appear to leak through the net of postulated phonetic laws, there should be an explanation for them. In the words of Karl Verner, "There must be a rule for irregularity; the problem is to find it." In searching for a rule that would explain certain residues of Grimm's Law, Verner found it necessary to go beyond the segmental environment into the accentual systems of Indo-European and Germanic, though the condition in the rule he discovered is still phonetic. In recent years, an impressive body of evidence has accumulated that there are diachronic rules which depend on conditions that are altogether non-phonetic, e.g. factors which are morphological and syntactic. That this is so should not be surprising: in phonology, diachronic rules frequently leave counterparts in the form of synchronic rules, and there are numerous instances where synchronic rules are obviously dependent on 'grammatical prerequisites.'

Furthermore, the stock of morphemes in a language is often partitioned into several layers according to non-phonetic criteria, where these layers exhibit different phonological behavior. To a large extent, such partitions correlate with the historical sources of the various layers, e.g. Romance versus native morphemes in English, or Chinese versus native morphemes in Japanese.

But what about the residual forms which remain, after we have taken into account the phonetic and morphological factors and the multilayered structure of the vocabulary? What explanation can we give for sounds changing differently under completely comparable conditions? Here the usual points about borrowing, dialect mixture, analogy, homonym prevention, the effects of tabu and phonetic symbolism, and functional load suggest themselves. Contrasted with the sweeping scope of phonetic laws, which have direct or indirect physiological motivation when internally induced, these suggestions appear unsatisfyingly ancillary and particularistic. In some cases, such

suggestions are completely ad hoc and unconvincing, but they remain in the literature mainly for lack of alternative explanations.

In this study we will examine the various dimensions along which phonological change is implemented: chronological, lexical, and phonetic. Many types of change are phonologically abrupt and require long spans of time to diffuse across the lexicon, as will be discussed in detail below. Since living languages are constantly undergoing change, we should expect to find many seeming exceptions to changes which have not completed their course. These forms are not true residues, in one sense, even though we cannot specify them in general terms either phonetically or morphologically, since in time the appropriate phonological changes will reach them and make them regular. True residue may result, however, as a consequence of the chronological intersection of competing changes. To this extent the neogrammarmarian doctrine must be modified. A sound change is regular if no other changes compete against it. But there are situations in which two (or more?) changes are applicable to the same subset of morphemes at the same time. Such situations leave residues which are the direct consequences of sound changes that were prevented from running their full course.

1.1 Relative Chronology

To develop this point we need to make two assumptions. First, we must assume that sound changes take varying periods of time for their operation. For lack of precise information, let us say that a sound change may take anywhere from several years to many centuries for its operation. The longer a sound change takes, the more likely it is to encounter a competing sound change.

Two sound changes may be complementary in their periods of operation. Indeed, many authorities on sound change seem to consider only this type of time relation, at least, they do not mention other types. Thus, in speaking of "relative chronology," Bloomfield (368) refers to "the succession of changes," and Hoenigswald (1960:112) to "the question of possible formal relationships between successive sound changes" (emphasis added). In addition to successive (i.e. complementary) time relations, we must also consider sound changes in coincident, incorporating, and overlapping relations with each other. For convenience we will refer to these three non-complementary types of time relations as intersecting. Two sound changes are intersecting if and only if the period of operation of one is partly or wholly concurrent with the period of operation of the other.

Although it is extremely difficult to determine the exact chronological relations between two changes, it is possible to make some elementary inferences. Let us think of phonological changes as diachronic rules, parallel in formal structure to the synchronic rules in a grammar. Thus, consider two diachronic rules, R1 and R2, which are related in a sense to be discussed in §1.4. Suppose that R1 changes some element X to Y, and R2 changes Y to Z; suppose, further, that the two rules operate under the same condition C. If all X's in condition C have changed to Z, then we can infer that R1 preceded
1. Competing Changes as a Cause of Residue

R2. Or, if all Y's have changed to Z and all X's to Y, then R2 must have preceded R1. Since there is no residue, we conclude that R1 and R2 were virtually complementary.

The situation is more complex if, under some constant condition C, R1 changes X to Y and R2 changes Y to X, yielding a case of flip-flop. Here, if all X's and Y's have indeed exchanged identities, we may infer that R1 and R2 are virtually coincident — for had R1 extended beyond R2, some of the X's produced by R2 would have been changed back to Y's; and similarly for R2.

1.2 Flip-Flops

Flip-flops pose an interesting problem for rule formalism in phonological theory. It is easy to see that when the elements involved contain the same number of segments, R1 and R2 may be collapsed by using a binary variable. This possibility of collapsing them captures nicely the coincident nature of the two rules. In fact, these cases strengthen the argument for binary specifications for those features which participate in flip-flops, since a specification that ranges over three or more values will introduce obvious difficulties when we try to collapse rules like R1 and R2.

A problem arises, however, when X and Y do not contain the same number of segments. We may consider an example from some Min dialects of Chinese. In Amoy, the morphemes for 'fire', 'age', and 'akin' are pronounced with /(#)/; in Longxi these same morphemes are pronounced with /(C)e/. On the other hand, the morphemes for 'chicken', 'shoe', and 'plough' are pronounced with /(C)we/ in Amoy but /(C)e/ in Longxi. If we assume for the sake of the present discussion that these morphemes retain in one dialect (say, Amoy) the earlier forms from which the forms in the other dialect (say, Longxi) are derived, then Longxi is separated from Amoy by two rules. One rule change /e/ to /we/, the other /we/ to /e/. These rules are disjunctive in their application, for to apply them conjunctively in one order would produce only /e/ and in the other order would produce only /we/. This suggests the use of a variable. Yet the variable cannot be attached to any distinctive feature, since it is the entire segment /w/ that is involved. We need to extend the use of the variable in a basic way to apply to whole segments in order to be able to describe this particular type of flip-flop, as illustrated in the rule below. In such rules the specification '+' when attached to the feature 'segment' denotes the presence of the segment with all its features, while the specification '--' denotes its absence:

```
\[
\begin{array}{c}
\text{segment} \\
\text{syllabic} \\
\text{consonantal} \\
+\text{velar}
\end{array}
\] 
\rightarrow 
\begin{array}{c}
\text{segment} \\
\text{syllabic} \\
\text{consonantal} \\
+\text{velar}
\end{array}
\rightarrow 
\begin{array}{c}
+\text{syllabic} \\
-\text{consonantal} \\
+\text{palatal} \\
.\text{mid}
\end{array}
\]
```

The phonetic history of these morphemes needs to be worked out in detail to see if the
assumption we made is true. It may be the case that the two dialects developed along altogether different lines. Or /æ/ may have been /o/ (i.e. a mid palatal labial vowel) at the time of the flip-flop, which would permit the more usual type of rule collapsing. In any case, it is at least possible that we have here two coincident rules which require us to extend the way in which variables are used in phonology. This extension is of great interest for our understanding of phonological theory.9

1.3 Lexical Diffusion

The second of the two assumptions referred to in §1.1 has to do with the way in which a change operates within a given period of time. Suppose that the change is from sound X to sound Y. Without concerning ourselves with the intriguing question of how the change came into the language in the first place,10 let us consider how it might actually implement itself. Ideally, before the change, all speakers will use sound X in all relevant morphemes; after the change, all speakers will use sound Y in the same set of morphemes. The dimension of time may be studied in each of three relatively independent parameters: (1) phonetic, i.e. from sound X to sound Y; (2) lexical, i.e. from morpheme to morpheme in the relevant part of an individual’s vocabulary; and (3) social, i.e. from speaker to speaker in the same dialect.11

With respect to the first of these parameters, we have in mind the familiar controversy of whether the change from X to Y is gradual or abrupt — i.e., whether sound changes occur always by successive slides, or always by abrupt leaps, or by both mechanisms, depending on the phonetic character of the sounds involved. By the term ‘gradual’ we refer to the ‘imperceptible increments’ type of shift as suggested, for instance, by Jespersen’s metaphor of sawing logs, in which, in the absence of a standard measure, the errors may have a cumulative effect. Jespersen’s metaphor has recently been quoted with approval by Hockett, who thinks that changes occur “mostly on a much finer-grained scale” still, and that they are “totally out of awareness” of the speaker.12

It is obvious that the gradual view of phonological change cannot be correct for many types of sound change, such as (1) when X and Y involve different articulators between which there is no physiological continuum (and it is implausible to imagine that there is any meaningful acoustical continuum; in the f > k in Hawaiian, for example, there is no evidence whatever for a phonetically intermediate stage); (2) flip-flop and metathesis, which are paradigmatic and syntagmatic variants of each other, and where the gradualist would be forced to shunt one of the elements off the collision course onto an artificial detour; (3) voicing, nasalization, and related processes, where our control of the relevant articulators is quite gross; (4) segment addition or deletion, since many sound types are either present or absent, but never present in gradient quantities. Furthermore, there is good reason to believe that many types of sound change must be regarded as operating at a phonological level that is much more abstract than the phonetic level, and therefore phonetically non-gradual. So, for a word like ‘acclimatize’ in which the pronunciation changes from [sklæmit ], the only pronunciation found in some
older dictionaries, to [ækɪmɛj], where all three vowels are different (in addition to the
difference in accent pattern), it is surely unrealistic to suppose that there was a gradual
and proportionate shift along all four phonetic dimensions. Clearly, the way to explain
such a change is to posit a diachronic rule of accent fronting which is followed by the
same body of stress adjustment and vowel-reduction rules which apply to a large sector
of other English words.

Even more obvious than the points made above is the fact that nonphonic changes
(e.g. lexical, morphological, and syntactic) cannot be accommodated at all within
the gradual view. In fact, it remains to be shown that the gradual view is applicable to
any sound change at all, even to those cases where there is some degree of surface
plausibility, i.e. where X and Y differ along continuous articulatory dimensions, as with
vowels and tones, over which significant scattering is observed. As Hoenigswald (73)
remarked, "The doctrine of gradual phonctic change may turn out to be remnant from
pre-phonemic days," when the multi-dimensional continua of speech had not yet been
successfully quantized.

Without pursuing this topic further, it is sufficient for the argument here to grant that
at least a large class of sound changes must be implemented abruptly, i.e. must go
directly from X to Y without passing through minute intermediate stages. This is to say
that the phonetic difference between X and Y is sufficiently gross to be potentially
phonemic, and a normal speaker of the language can distinguish between X and Y easily
and consistently without any special training.

The process of diffusion within a speaker’s vocabulary may likewise be thought of
as being either abrupt (i.e. all relevant morphemes change ‘simultaneously’) or gradual
(i.e. the change affects the relevant morphemes severally in succession). There are
then four logical possibilities in viewing how a sound change operates on an individual’s
vocabulary:

1. phonetically abrupt and lexically abrupt
2. phonetically abrupt and lexically gradual
3. phonetically gradual and lexically abrupt
4. phonetically gradual and lexically gradual

Of the two possibilities which involve a phonetically abrupt implementation, (1) is refuted
by elementary observations: it implies that all morphemes with only X-pronunciations for
a given speaker would suddenly have only Y-pronunciations, which is obviously unac-
tceptable. A sound change takes time, not only for the collective vocabulary of a speech
community, but for the vocabulary of individuals as well. (2) is the view taken by the
neogrammarians and continues to be widely accepted today. The hypothesis of lexical
diffusion stipulates (2) and (4). Of these two, (2) is the more compelling: for given that
the phonetic implementation is abrupt, and that an individual’s vocabulary does not
change all that suddenly, the obvious conclusion is that what actually takes place is a
kind of diffusion from morpheme to morpheme in his vocabulary. This diffusion within a
lexicon is basically the same mechanism as the more observed forms of diffusion across
dialects or languages, and diffuses only in its scope of operation; lexical diffusion is more local, the other forms are more global.

We do not need to insist that lexical diffusion is the only means by which the pronunciation of morphemes changes. It is sufficient for the argument here that this is one of the primary means through which a sound change implements itself. According to this view, during the early phase of the change only a small sector of the relevant morphemes is affected. Some of the affected morphemes may change to the X- pronunciation directly. Other morphemes, however, will at first have both the X- pronunciation and the Y- pronunciation, fluctuating either randomly or according to some such factor as tempo style. (For the most part, morphemes do not have more than two pronunciations. In the phonetic literature, these dual forms have sometimes been referred to as 'doublets'.) But the X- pronunciation will gradually be suppressed in favor of the Y- pronunciation. These doublets, then, serve as a kind of psychological bridge between the two end-points of a sound change, carrying along with them even those morphemes which do not go through a doublet stage.

This hypothesis of lexical diffusion suggests that, at any given time in any living language, we should expect to find several sets of morphemes with dual pronunciations. Such examples are not difficult to obtain, although they are not abundantly reported in the literature, presumably because perception is a function of expectation; and their existence is not suggested by the dominant theory of sound change illustrated earlier in the works of Jespersen and Hockett. Thus, in Chinese dialects there are large sectors of morphemes which have two pronunciations, one 'literary,' the other 'colloquial' (cf. Peking University 1962). For English, the pages of any good pronouncing dictionary (e.g. Kenyon & Knott 1944) show that many morphemes have two pronunciations, such as those involving accent pattern [əbdəmən / abdəmən], postvocalic r [səprǎj / sapərəj], vowel labialization [kætalək / kætalək], vowel length [ruː / ruː], syllabicity [təwl / təwel], j-glides [nu / nú], voicing of intervocalic obstruent clusters [ɛksit / ɛgət], and so on. Dialect geography can tell us which areas favor which clusters of pronunciation. In actual fact, of course, many of the dual pronunciations are used by the same speaker.

As Vogt remarked (1954:367): 'At any moment, between the initiation and the conclusion of these changes, we have a state characterized by the presence of more or less free variants, so that the speakers have the choice between alternative expressions. In each case the choice will be determined by an interplay of factors, some linguistic, some esthetic and social, an interplay so complex that most often the choice will appear as being due to pure chance.... What therefore in a history of linguistic system appears as a change will in a synchronic description appear as a more or less free variation between forms of expression, equally admissible within the system.' Recent advances in sociolinguistics hold the promise that perhaps some of the social factors determining linguistic choices of this kind can be investigated in a roughly quantitative way.

It is not always easy to determine if a given situation is the result of dialect borrowing, or if it is due to the process of lexical diffusion as described here. To say that dialect D₁ borrowed a particular set of Y-pronunciations from dialect D₂ can only be par-
tial answer at best. It merely tells us that the phonological explanation for the X-to-Y change does not lie in D1. We must trace back to the dialect in which the change originated, if the question ‘why did X change into Y?’ is to be fully answered. In those cases where we are fortunate enough to catch the inception of a change at one or two isolated sites, influence from neighboring dialects can be ruled out and we can have a clearer view of how the change implements itself. Thus when a dialect dictionary (Peking University 1962) shows that there is an ongoing change from the retroflex series to the labiodental series in Xian which is not reported in any of the neighboring dialects, we can hope that a detailed historical study of the Xian dialect will reveal the workings of lexical diffusion.

As the change continues to operate, an increasing portion of the relevant morphemes will become affected. Cases have been reported where, for extralinguistic reasons, a sound change ceased (or even reversed itself) after it had operated over only a part of the relevant vocabulary. It is important to note that such cases (if they are real) are in direct conflict with the almost universally accepted view of how phonemic split can come about, as summarized by Moulton (1937:1994 f.). As Harms has recently put it (1967): "any claim for a split... must be supported by the environmental conditioning which gave rise to the putative split" (166) and "no other basis for split consistent with the synchronic structure of language is known" (169). If cases of incomplete changes can be established with certainty, then our understanding of the basis of phonemic split and, more broadly, of the regularity hypothesis itself must be revised.

If, however, the change runs its full course (which seems to be frequently the case), then all the X-pronunciations will be replaced by Y-pronunciations — unless another sound change comes into the picture to compete for some of the same morphemes relevant to the X-to-Y change. Notice that the view of sound change as competition between co-existing rivals is less constraining on the degree of phonetic similarity between the two pronunciations than one which construes one pronunciation as some sort of direct and gradual phonetic mutation of the other. There is thus no a priori reason to believe, for instance, that sound changes necessarily affect only one distinctive feature at a time. A fortiori, I see no basis at all for Sweet’s assertion (1888:15) that “Such a change as the frequent one of [li] into [ai] presupposes a number of intermediate stages: [i], [li] [el], [el], [el], [ai], etc. Hence also there are no simultaneous changes of a sound, only successive ones. Thus we cannot suppose a simultaneous opening and unvoicing of [m], but only such series as [m], [b], [b], [d]." Nevertheless, the view of sound change exemplified by Sweet here may be actually the dominant view of this area of linguistic scholarship. The notion of doublets as an intermediate stage in sound change was specifically rejected by Wheeler (1887:5), who wrote: “It is to be noticed that the operation of the laws of sound is unconscious and gradual, so that the old form cannot, except through mixing of dialects, survive alongside the new” (emphasis added).

Given the picture roughly sketched out in this section, sound changes can be seen to proceed very much along the same principles as other, non-phonetic linguistic chan-
ges. A sound change may involve, simultaneously, several phonetic dimensions which are related to each other by phonological rules, as illustrated by the word 'acclimate' above, though it is typically simple and physiologically motivated. Over a half-century ago, Sturtevant described the situation quite well, though he did not use the term 'lexical diffusion,' when he wrote the following passage: 'We have seen that many sound-changes are irregular when they first appear and gradually become more and more regular. The reason is that each person who substitutes the new sound for the old in his own pronunciation tends to carry it into new words. The two processes of spread from word to word and spread from speaker to speaker progress side by side until the new sound has extended to all the words of the language which contained the old sound in the same surroundings' (82). In addition, he noted that "Such a spread of a sound-change from word to word closely resembles analogical change; the chief difference is that in analogical change the association groups are based upon meaning, while in this case the groups are based upon form" (80).

Lexical diffusion may be schematized as in Table 1, which illustrates the basic paradigm of phonological change. At the beginning of the time span, \( t_1 \), segment \( A \) occurs in four distinct contexts, \( C_1 \), \( C_2 \), \( C_3 \), and \( C_4 \). These contexts may be specified in either phonological or morphological terms. At \( t_2 \), \( A \) has become \( B \) in the context \( C_1 \), creating an alternation in the sound system. We may think of \( C_1 \) as the primary context, as it is most likely the case that \( A \) is changed to \( B \) through an assimilatory process motivated by \( C_1 \).

<table>
<thead>
<tr>
<th>Table 1. Basic Paradigm of Phonological Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>( C_1 )</td>
</tr>
<tr>
<td>( C_2 )</td>
</tr>
<tr>
<td>( C_3 )</td>
</tr>
<tr>
<td>( C_4 )</td>
</tr>
</tbody>
</table>

All the morphemes which have \( A \) in the context \( C_1 \) will not change at the same time, of course. Some of them may even lag and change only after \( t_3 \), or maybe even later. It would be of interest to determine if such lags typically do or do not occur; and if they do, then to what extent. The table shows that by \( t_5 \) all of the \( A \)'s have changed into \( B \)'s. That is, the conditions for the change have relaxed until finally the change has become unconditioned. Formally stated, this can be seen as a process of successive simplification in the diachronic rules that will ultimately have the effect of eliminating whatever alternation was caused by the primary contest. In languages like Chinese, which have little morphophonemic alternation, since they have little morphological inflection, the effect of change upon the morphophonemic representations is correspondingly more direct. Unless we can reconstruct the history of this diffusion and determine what was the primary contest, the change of \( A \) to \( B \) in the non-primary contexts will seem to lack phonetic motivation, or even to be counter-phonetic.
1.4 Rule Relations

The two assumptions we have made above allow us to conceive the following situation. Suppose we have two diachronic rules which intersect in time. In the schema below, I, O, and C denote input, output and condition; the subscripts denote the number of the rule.

\[
\begin{align*}
R_1 & : I_1 \rightarrow O_1 / C_1 \\
R_2 & : I_2 \rightarrow O_2 / O_2
\end{align*}
\]

Suppose, also, that \( R_1 \) and \( R_2 \) operate across the vocabulary in the manner of lexical diffusion suggested above. One further condition must be satisfied before these two rules can cause residue: they must be in competition with each other.

The competing relation in diachronic rules is approximately the counterpart of the ordering relation in synchronic rules. Two synchronic rules are ordered with respect to each other, if, for an arbitrary input set, the output set differs according to the sequence in which the two rules are applied. Diachronically speaking, two rules are in competing relation if there are morphemes whose phonetic histories would differ according to the sequence in which the two rules are applied. Since ordering is transitive, two rules may be indirectly ordered via intermediate rules. From recent studies we have learned that relations among phonological rules may be quite complex: thus, rules which assign specifications to prosodic features for large syntactic constructions may need to be ordered cyclically. There are rules which 'persist' in that they apply whenever applicable,\(^{18}\) such rules usually have strong phonetic motivations. Rules may be conjunctive or disjunctive. Furthermore, there may be cases where a rule may either precede or follow a block of rules but must not apply within the block.\(^{19}\) In the present discussion, however, we will consider only the simplest case: the relation between two rules.

The relation between a pair of rules may be studied by reference to the schema given above for \( R_1 \) and \( R_2 \). Recall that, in its formal representation, each of the six parts in the two rules is a matrix of feature specifications. Thus, given two matrices, \( M_1 \) and \( M_2 \), neither of which is an empty matrix, \( M_1 \) may contain \( M_2 \) (\( M_1 ⊆ M_2 \)), \( M_1 \) may be contained in \( M_2 \) (\( M_1 \cap M_2 \)), or \( M_1 \) may equal \( M_2 \) (\( M_1 = M_2 \)). The assertion \( M_1 ⊆ M_2 \) is of course equivalent to the assertion \( M_2 \cap M_1 \). We use the formula \( M_1 ⊆ M_2 \) to denote that either \( M_1 \cap M_2 \) or \( M_1 = M_2 \); and similarly for \( M_1 \cap M_2 \). The formula \( M_1 = M_2 \) denies both \( M_1 ⊆ M_2 \) and \( M_1 \cap M_2 \). On the other hand, \( M_1 \cap M_2 \) means either \( M_1 ⊆ M_2 \) or \( M_1 \cap M_2 \).

The reader will note that if \( M_1 ⊆ M_2 \), then \( M_2 \) actually refers to a larger class of phonological situations. A matrix that represents high front vowels will contain a matrix that represents high vowels, though the latter of course refers to a more general class of segments. The formalism does not cover cases where \( M_1 \) and \( M_2 \) overlap but neither contains the other. Such a situation would arise if \( M_1 \) represented high front vowels and
M_2 represented a sequence of two high vowels.

In terms of the formalism introduced above, it is easy to enumerate all the possible relations R_1 may have to R_2. Thus, we see that the matrix O_1, for instance, may participate in any of the following relationships. If O_1 = ø, clearly R_1 is a deletion rule.

If O_1 ⊆ I_2, then we have a number of subcases. If C_1 = ø, C_2 = ø, or C_1 ∩ C_2, then R_2 will apply to O_1, as example 1 shows:

\[(1) \quad R_1 : \text{mid vowels} \quad \rightarrow \quad \text{high vowels} / \ldots / C \\
R_2 : V \quad \rightarrow \quad \ddot{V} \quad / \ldots / \text{nasals} \]

If we impose the further condition that I_1 = I_2, we have what has been called a feeding relation between the two rules (example 2):

\[(2) \quad R_1 : s \rightarrow z / \ldots \dot{V} \\
R_2 : z \rightarrow r / \ldots \ddot{V} \]

In example 3, where I_1 = I_2 and C_1 ⊆ C_2, we have a bleeding relation: \(^{20}\)

\[(3) \quad R_1 : t \rightarrow \ddot{c} / \ldots \dot{i} \\
R_2 : t \rightarrow d / \ldots \ddot{V} \]

In example 4, where I_1 = O_2 and C_1 = C_2, we have an instance of what might be called a replenishing relation.

\[(4) \quad R_1 : n \rightarrow \ddot{n} / \ldots \dot{V} \\
R_2 : l \rightarrow n / \ldots \ddot{V} \]

Example 5 illustrates a voiding relation, where C_1 = I_2.

\[(5) \quad R_1 : a \rightarrow o / \ldots l \\
R_2 : l \rightarrow n \]

It should go without saying that there are rules which are not in a competing relation, as in example 6.
The four relations illustrated by examples 2-5 have interesting relations within themselves. Replenishing and feeding, for example, are partial inverses of each other. If a pair of ordered rules includes a flip-flop rule, then the relation between the two rules is both feeding and bleeding. (In fact, it is possible to obtain a simultaneous feeding and bleeding relation even when neither in the pair is a flip-flop rule.) In addition, there must be dozens of other possible relations that can be uncovered in an exhaustive study that exploits more fully the formalism introduced here. Two sound changes must have a competing relation to cause a residue; we must be satisfied here with merely illustrating some of the more obvious competing relations.

1.5 Discussion

The conjecture offered here is that residue may be caused by two competing sound changes that intersect in time. A definitive proof of this conjecture must await case studies with careful and detailed documentation. My purpose here is merely to give it an airing, in the hope that scholars may be interested in investigating its validity in their respective areas of expertise.

If the picture of sound changes sketched above is valid, then the effects that they have on the phonological structure of the lexicon are certainly much more complicated than the neat universe suggested by the ‘regularity hypothesis’. Given that R₁, R₂, and R₃ operated in complementary periods, then the class of relevant morphemes can at most be partitioned into 2³ subsets, where ‘+’ indicates that the rule has been applied and ‘−’ indicates that it has not, as shown in Table 2. Only subset I above is the perfectly regular class; subset VIII has somehow escaped all three sound changes.

If the three sound changes overlap, then the number of logical possibilities is significantly increased. The morphemes of subset I may be further divided according to the six possible orders in which the three rules may apply. Subsets II, III, and IV each further divides according to two possible orders. In fact, the very notion of ‘residue’ becomes quite subtle as it becomes increasingly difficult to determine which one of these numerous subsets of morphemes followed a ‘regular’ development.

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>VI</th>
<th>VII</th>
<th>VIII</th>
</tr>
</thead>
<tbody>
<tr>
<td>R₁</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>R₂</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>R₃</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
</tr>
</tbody>
</table>
This view of diachronic phonology has obvious implications for synchronic analysis. Current theories more or less assume that all morphemes uniformly go through the same full body of phonological rules, those which exhibit alternation as well as those which do not. It is clear from the diagram above, however, that there may be a considerable amount of cross-classification as to which rules are applicable to which morphemes, so that the neat picture of all-rules-apply-to-all-morphemes becomes highly suspect. We need to make much more use of rule features in synchronic phonology in order to capture the diversity in the lexicon that is due to the various cross-currents of sound change.

Some of the standard concepts of diachronic phonology will need re-examination. It is generally believed, for example, that splits can result only from a conditioned change, and that contrasts are possible only after something happens to the condition of the change. But if we accept the fact that a sound change (conditioned or unconditioned) may not complete its course due to other competing changes, then clearly we may also need to recognize incomplete sound changes as a cause of splits.

The conjecture presented here can only be invalidated by demonstrating that all (or at least competing) sound changes are complementary in their periods of operation, or that lexical diffusion is an untenable view of how sound changes spread. Both the assumption of relative chronology and the assumption of lexical diffusion, as they are sketched above, seem to me, however, to be eminently reasonable in the light of our present state of knowledge. Therefore, I would think it extremely unlikely that the conjecture will be invalidated on these grounds. Also, it should go without saying that the conjecture cannot be falsified by demonstrating that certain phonetic laws do not have residues or that certain residues are caused by other factors. (We cannot prove that the platypus does not lay eggs with photographs showing a platypus not laying eggs.)

Probably it is not much easier to validate the conjecture with an incontrovertible proof, primarily because such subtle factors are involved. But there are several considerations which give it indirect support. Although much more attention is usually given to the regular aspects of sound changes than to their residue, a closer scrutiny reveals that even the best phonetic laws are frequently ridden with irregularities. 21 When all currently available explanations fail, we must examine our premises and be prepared to accept a concept of sound change that is more detailed and more complex.

One of the most striking achievements of modern linguistics is the wealth of reliable information gathered about the phonetics of a great diversity of languages, on the basis of which we can make some preliminary historical guesses. The evidence is overwhelming that almost all sound changes come from an extremely small common inventory. 22 Many of these changes are now beginning to find explanations in laboratory phonetics. To the extent that they are phonetically motivated, they are unidirectional or asymmetrical. The degree of parallelism in phonological histories is such that every argument for a common history of dialects on the basis of a common sound change is suspect unless it is buttressed by additional independent considerations. 23 In fact, this small inventory of changes (e.g. umlaut, vowel nasalization before nasals, tone lowering
with voiced consonants, palatalization and sibilation before grooved vowels, etc.) constitutes the empirical foundation for much recent research on marking conventions in phonological theory, especially the use of these conventions as linking rules.

The observation that sound changes are few in kind suggests that, given any two sound changes, there is a significant possibility that they are formally in competition with each other. This line of reasoning gives indirect support to the conjecture that residue may be caused by two competing sound changes that intersect in time.

Another consideration has to do with time. The longer sound changes take, the more likely it is that they will intersect, compete, and leave a residue. To measure exactly how long a sound change took, we would need to be able to determine the endpoints of its period of operation, which is nearly impossible in our present state of knowledge. Part of the difficulty stems from the fact that a sound change may begin sporadically in the vocabulary (i.e. the class of morphemes it affects can only be defined by enumeration); it then consolidates itself and becomes regular but conditioned; and the conditions may eventually be simplified until, finally, it becomes an unconditioned change. The phonetic condition that originally stimulated the change may create a ‘snowball’ effect across the lexicon, so that the condition itself eventually becomes irrelevant. This means that, in a perfectly justifiable sense, one could argue that most conditioned changes, even though they have no residue, are really subchanges of larger changes which have not completed their course. If we accept such arguments, then we must be prepared to recognize sound changes as typically endowed with a much greater longevity than is commonly thought.

As an interesting example, let us consider some English rules of vowel shortening which have been recently discussed by Kiparsky. In Old English, vowels were shortened either before three consonants or before two consonants and two more syllables within the same word. Beginning in Early Middle English, the condition of this rule was simplified to either before two consonants or before one consonant and two more syllables within the same word. The latter rule is still with us, as evidenced by such morphophonemic alternations as in sleep/ slept and felina/felinity. If we take these rules to be synchronic manifestations of an underlying sound change, then we see that the deletion of a single consonant in the condition of the change has operated over roughly a thousand years! Such examples, of course, are not hard to find. Another sound change from the history of English is the loss of /h/ in presonorant position, which still has residual forms in many dialects despite many centuries of operation.

Surely, if we give more consideration to the dimension of time, much of the 'unstructuredness' observed by dialect geographers and in sociolinguistic studies of language usage can be better reconciled with the necessary faith that our linguistic behavior is lawful. Commenting on the general principles of biological evolution, Huxley (1953:28) noted that in synchronic descriptions "we freeze the process into a set of unreal static pictures.... Only in the longest perspective... do the overall processes of evolution become visible." Within the narrow time perspectives available to us, numerous competing cross-currents appear to be vying for the phonological future of
every word in the lexicon. Without adequate time-depth, it is often difficult for us to know which changes are sporadic, which will persist, which come from past millennia, which are receding, and which are just coming into the language. To understand such a complex dynamic situation, we must attempt to isolate for careful analysis each of the various interwoven factors — the physiological, the structural, the societal, and yet others. From the factors examined here, it seems that the 'regularity hypothesis' must be modified to allow for residue caused by competing sound changes which intersect in time.27

Notes

1. That there should be non-phonetic conditions for sound change was expressed in this way by Sapir (1921:183 f.): "Every linguist knows that phonetic change is frequently followed by morphological rearrangements, but he is apt to assume that morphology exercises little or no influence on the course of phonetic history. I am inclined to believe that our present tendency to isolate phonetics and grammar as mutually irrelevant linguistic provinces is unfortunate. There are likely to be fundamental relations between them and their respective histories that we do not yet fully grasp." It should be noted that Sapir's view is not compatible with that of the neogrammarians, who "define sound-change as a purely phonetic process... that affects a phoneme or a type of phoneme either universally or under strictly phonetic conditions" (Bloomefeld 1933:364). Non-phonetically conditioned sound changes are discussed at length in Postal 1968. In several recent works by Makkai (e.g. 1968) we also find carefully documented histories of how morphological structures influence sound change.

2. Cf. the following passage from Bloomfield (364): "We must suppose that, no matter how minute and accurate our observation, we should always find deviant forms, because, from the very outset of a sound-change, and during its entire course, and after it is over, the forms of the language are subject to the incessant working of other factors of change, such as, especially, borrowing and analogic combinations of new complex forms" (emphasis added). Bloomfield did not consider competing sound changes among the factors responsible for residue; and, as far as I know, neither has anyone else.

3. I have attempted to provide a statistical basis for the notion of functional load in Wang (1967a). Results so far based on the measure developed in that article, or on various modifications of it, have led me to be quite pessimistic as to the value of this notion for our understanding of the causes of sound change. There is some surface plausibility to the suggestion that contrasts which distinguish larger numbers of higher frequency words are more resistant to loss. When we recall, however, the tremendous amount of various types of redundancy that is typically present in the speech situation, it appears that individual contrasts are not really all that crucial for effective communication. It is not difficult to find cases where contrasts have been lost on a grand scale: for example, the number of distinct syllables has been reduced by a factor of 3:1 from Middle Chinese to Modern Pahinise (cf. also the results of King 1969).

4. These four terms are adopted from Wells (1949). They were used for still another purpose in Block (1953).

5. I have elsewhere discussed a case of tonal flip-flop in a Min dialect of Chinese (Wang 1967b:102-3).

6. Variable notation was introduced by Halle (1962).

7. Binarily specified features can be used to describe alternations among three or four elements (as with vowels or tones) by attaching paired variables to different features; cf. Wang (1968).

8. These examples are taken from Tung (1960:997-1010). Robert L-W. Cheng has called my attention to a similar situation with /æ/ and /æ/ in Northern and Southern Taiwanese.

9. Note that in this formalism the deletion of /æ/ has a much more conventional interpretation than the addition of it. For the latter we have to assume that there is an implicit /æ/ in the underlying form which surfaces when the specification of the feature segment is changed.

10. Another use for the variable notation has been recently proposed by W. Labov to investigate statistical regularities in the application of rules. Later sections of this paper should make it clear that Labov's hypothesis on rule application and the hypothesis of lexical diffusion are related but distinct notions. Only in the latter hypothesis is it claimed that rule applicability in many cases must be specified via a via individual items in the lexicon.

11. This question, as well as many related issues, is discussed with admirable balance and clarity in Weinreich, Labov, & Herzog (1968). Postal has recently expressed the rather extreme view of phonological
change (p. 283 and elsewhere) in which all changes are stylistic and non-functional, and phonetics plays no role in considerations of sound change. It is difficult to see how such a view can be justified in the light of the abundance of different types of sound change which can easily be shown to have phonetic motivation, as against the paucity of attested 'counter-phonetic' changes. In fact, many changes are only apparently counter-phonetic since it can be shown that they are triggered by other changes in the system which are phonetically motivated. It may very well turn out upon deeper examination that all internal phonological changes are directly or indirectly actuated by phonetic factors, as envisioned in Whitney (1877).

11. We preserve the convenient illusion of 'dialect' here for ease of discussion, realizing that isoglosses intersect in diverse fashions and that the concept is often no more than an idealization. Much of what we say here about lexical diffusion is true across dialects as well as within dialects.

12. See Hockett (1955:202). For the cases where errors (or imprecision of pronunciation) build into a sound change, neither Jespersen nor Hockett explain why these errors should be cumulative in effect rather than mutually cancelling, or why the overwhelming majority of sound changes are unidirectional - i.e., why the relation between X and Y is anti-symmetric. The claim of sound changes occurring 'totally out of awareness' is based on the neogrammarians' belief that if the variation were noticeable, it would be corrected and sound change would not take place. Recent studies in sociolinguistics show, however, that speakers are very much aware of many types of subphonemic fluctuation which are socially marked; see especially Labov (1968).

13. For a relatively early statement in favor of the 'abrupt leap' view, see Sommerfelt (1923). Still earlier, Sturtavant (1917:73) noted that "Even some changes which might take place by imperceptible stages are nevertheless observed to involve at certain times an easily perceptible variation between words or between speakers." Compare these opinions with that of Sweet (1888): "The most serious of my defects of method was my rejection of the principles of gradual sound-change in favour of change per saltum..." (vii); and "... all sound-change is gradual: there are no sudden leaps in the phonetic history of a language" (15). For his theoretical position, Sweet claims to "owe most to Paul and Siervers" (xii).

14. We refrain at this point from speculating on certain fascinating questions: why some morphemes change earlier than others, whether morphemes follow the same change schedule from speaker to speaker, and how speakers influence each other's change schedules. These questions are probably in large part insoluble on the basis of purely linguistic considerations, since they appear to depend on many social factors. At any rate, the answer can not be as facile as Sturtavant pictured it when he wrote (83): "... it seems safe to say that the likelihood that a sound-change will become regular varies directly as the number and the frequency of the words which induce it." Opinions similar to Sturtavant's have been expressed more recently by Greenberg et al. (1954:146) (see also fn. 3 in the present paper for further discussion). Current work on sound change at a more microcosmic level (e.g., Labov) may eventually lead us to a better understanding of these questions, where consideration is given to the social aspects of language use, as well as to the structural aspects internal to the language.

15. I am indebted to George Grace for drawing my attention to this clearly worded statement.

16. Sturtavant (76) gives an example of an unsuccessful sound change: "In Latin there was at one time a tendency to lengthen short vowels before gn... but... the pronunciation with a short vowel finally prevailed. In this case the net result of the incipient change was to leave things as they were at first." When the morphemes affected by an unsuccessful sound change do not get changed back, we have two classes of forms. Whether the new or the old forms should be considered regular is an interesting terminological question that depends in part on how unsuccessful the second change was. Examples of unsuccessful sound changes, which are noted as such, are scarce in the literature; of closely related interest here are cases that have been reported where Y changes back to X, which has been termed 'retrograde' (Weinreich 1958).

17. I have replaced Sweet's Visible-Speech symbols with a more familiar phonetic notation. The theory that sound changes occur by single features is especially unacceptable with respect to ephenthesis and deletion. For example, in Halkow Chinese an /ŋ/ has developed after /m/. A theory that requires us to assign relative chronology to the various features of the epenthetic /ŋ/ is clearly misguided; and similarly for cases when segments are lost.

18. For further discussion on ordering relations among phonological rules, especially regarding persisting rules, see Chafe (1967). For general theory, see also Postal (1968), and Chomsky & Halle (1968).

19. The possibility of such cases was first suggested by C. Douglas Johnson in discussion (see Johnson 1967).

20. The terms 'feeding' and 'bleeding' were proposed by Kiparsky (1968).

21. Thus in Romance linguistics, a field that serves as a precious laboratory for historical methods, Bonfante writes (1947:345): 'It would in fact be easy to show that for many of the so-called phonetic
22. Without a detailed specification of how broad a range of phenomena each sound change should encompass, it would be meaningless to offer an exact number at this time.

23. Given dialects $D_1$ and $D_2$ and a sound change $R$ which began at time $T$, we can argue with force that $D_1$ and $D_2$ must have split from each other at least since $T$, because they do not share $R$. But the other side of the argument is considerably weaker: that $D_1$ and $D_2$ must not yet have split at $T$ because they share $R$. Languages, as well as dialects, can independently undergo the same sound change, with a probability that is directly correlated with the degree of similarity between the phonetic structures, but not necessarily related to the genetic affiliations involved.

24. Palatalization and sibilation are two very different phonetic processes, though they frequently go together, and both are predominantly regressive, as are most assimilations. The former is an assimilation to the tongue position of a vowel such as [i], which is articulated with the vocal tract narrowest at the palatal region, as in key and tea. The latter is an assimilation to the tongue shape of the vowel. High front vowels are typically articulated with a groove in the center of the tongue, i.e. with the center down and the edges turned up against the lateral teeth. The presence of such a groove is the distinguishing property of sibilants, whereby the air stream is channeled against the upper edge of the lower incisors. So the English changes of [k] and [t] to [s] as in the letter 'c' and in lunacy need not have gone through a palatal stage at all.

25. The statement by Greenberg et al. (149) that "in general, conditioned change is the diachronic aspect of the synchronic problem of conditioned allophonic variation" is, of course, quite incomplete. Morphophonemic variation also may reflect sound changes. Due to a variety of factors, synchronic rules may not faithfully reflect the phonological history. Thus, a sequence of diachronic rules may be telescoped into a single synchronic rule; i.e. $A \rightarrow B$, $B \rightarrow C$, and $C \rightarrow D$ go into just $A \rightarrow D$ if there is no synchronic motivation for position $B$ and $C$ as intermediate stages. The English vowel shift may be such a case; cf. Wang (1968). Or the effects of a diachronic rule of the form $A \rightarrow B/C$ may be more simply accounted for synchronically by a reversed rule $B \rightarrow A/D$. For further discussion on the relation between synchronic and diachronic phonology, see Greenberg (1959) and Zep (1965).

26. The idea is expressed so well that the full paragraph should be quoted here: "When we take an instantaneous snapshot, we freeze the process into a set of unreal static pictures. What we need is the equivalent of a film. We all know how a film record can be speeded up to reveal processes that are hidden from ordinary view — the dancing movements of a growing twig, the adventurous transformations of a developing egg. The same applies to our moving picture of evolution. If this is run at what seems natural speed, we see only individual lives and deaths. But when, with the aid of our scientific knowledge and our imagination, we alter the time scale of our vision, new processes become apparent. With a hundredfold speeding up, individual lives become merged in the formation and transformation of species. With our film speeded up perhaps ten thousand times single species disappear, and group radiations are revealed; we see an original type, seized by a ferment of activity, splitting up and transforming itself in many strange ways, but all the transformations eventually slowing down and stabilizing in specialized immobility. Only in the longest perspective, with a hundred-thousand-fold speed-up, do the over-all processes of evolution become visible — the replacement of old types by new, the emergence and gradual liberation of mind, the narrow and winding stairway of progress, and the steady advance of life up its steps of novelty!"

27. This study was supported in part by the National Science Foundation Grant GS1430, and grew out of group discussions conducted at the Phonology Laboratory in Berkeley during the summer of 1957. Detailed comments from M. Beeler, W. Chafe, C. D. Johnson, K. Zimmer, and C. J. Bailey have helped me improve the presentation of several points. I am also grateful for a summer of research at the Pacific and Asian Linguistics Institute of the University of Hawaii, where I had the benefit of extended discussions with George Grice.

Postscript 1977

Since the above discussion, a good amount of empirical work has been done within the framework of lexical diffusion. Instances of diffusion have been investigated in many
languages, including Chinese, English, German, Swedish. Much of this research is summarized in an article by Chen and Wang in *Language* 51, 255-81, 1975. In addition to language phylogeny, the concept of lexical diffusion has been applied in language ontogeny as well, in the work of Hsieh (Glossa 6, 89-104, 1972) and of Ferguson and Farwell (Language 51, 419-39, 1975).

It is clear by now that phonemes frequently are not the minimal units in historical change. Neither are they in the acquisition of phonological patterns. These observations highlight an intriguing issue which we may call the regrouping problem. If a phoneme is pronounced as several different sounds in a class of words, (and there are no phonetic conditions to relate the variants one to another), what are the mechanisms that guide the variants to regroup into the same sound again? In the case of acquisition, the regrouping is more understandable since adult models are readily available. But for historical change, especially for languages with no writing or with nonphonetic orthographies, the degree of regularity that is believed to exist is quite puzzling. Weinreich, Labov and Herzog, in their important synthesis of 1968, attribute the regrouping to "a rise in the level of social awareness of the change and the establishment of a social stereotype" (p. 167). But what prompts all the affected words to establish the same stereotype? Surely phonetic factors must play an important role as well in some of the cases; but which ones?

The regrouping problem has not been given sufficient attention because lexical abruptness is usually assumed. Even with variable rules of the sort introduced by Labov, the problem does not emerge with clarity because these rules do not distinguish individual lexical items, but average out their usage by statistical methods. Yet it seems to me that the solution of this problem is crucial to a proper understanding of the sound patterns of language.