A Dissertation on
Natural Phonology

David Stampe
CONTENTS

Preface vi
The Acquisition of Phonetic Representation vii
A Dissertation on Natural Phonology xxvi
   Chapter I: The Nature and Function of Processes 1
   Chapter II: The Organization of Process 43
   References 71
   Afterthoughts 75
Additional References 84

© 1979 David Stampe
All rights reserved

All volumes in this series are printed on acid-free,
250-year-life paper.
Printed in the United States of America
PREFACE

The present edition of my dissertation has been retyped, with minor corrections but without pagination changes, from the version submitted to the University of Chicago Department of Linguistics in December 1973. The latter was slightly abridged from a September 1972 version distributed under the title How I Spent my Summer Vacation.

The paper The Acquisition of Phonetic Representation, from 1969 Regional Meeting of the Chicago Linguistic Society, is reprinted here (again with minor corrections) as a convenience to the reader, because much of the discussion in the dissertation presupposes the paper.

The dissertation was originally without footnotes. I have added some here, as 'Afterthoughts,' mostly to explain some subsequent changes in the theory. These are not intended to be exhaustive. For details the reader may be directed to The Study of Natural Phonology, a 1979 paper by Donegan and me (see "Additional References") that furnishes a general survey of recent work.

Columbus, Ohio

July 1979
THE ACQUISITION OF PHONETIC REPRESENTATION*

Since 1965 I have been reporting to the Society on a theory of natural phonology, based on the assumption that the phonological system of a language is largely the residue of an innate system of phonological processes, revised in certain ways by linguistic experience. But since my previous talks have not been published, I have chosen this year to review some highlights of those talks rather than to report on more recent work.

A phonological process merges a potential phonological opposition into that member of the opposition which least tries the restrictions of the human speech capacity. Processes characteristically fall into contradictory sets, reflecting conflicting phonetic restrictions. Obstruents become voiceless irrespective of their context, because their oral constriction impedes the airflow required by voicing, while, on the other hand, they become voiced in voiced environments by "assimilation." Where these processes overlap, for example between vowels, there is a contradiction: an obstruent cannot be both voiceless and voiced. There are three ways by which such contradictions are resolved.

The most radical resolution is by suppression of one of the contradictory processes. The mastery of voiced obstruents in all

contexts entails the suppression of the devoicing process mentioned above, with a resulting opposition of voice in obstruents in all contexts except those in which the voicing process merges them.

The second resolution is by suppression of some part of a process—limiting the set of segments it applies to or the set of contexts it applies in. Implicit in each process are various subtle and strict hierarchies, ranging from the greatest generality which is phonetically motivated, to the complete suppression of the process. For example, the devoicing process may be limited to tense obstruents, but not to lax ones unless it is altogether suppressed, because tense obstruents are less favorable to voicing than lax ones. Or it may be limited to voiceless contexts, or initial and final positions, and so forth. Likewise the voicing process may be limited to contexts between vowels, between non-high vowels, or between low vowels, but it may not be limited to apply just between high vowels, for example. In this example the traditional parameter of "sonority" clearly plays a role. Besides such phonological limitations, there are also non-phonological limitations which may be imposed, but these are not relevant to this discussion.

The third resolution is by ordered application. Many languages lack a voicing opposition in obstruents (this can be ascribed to the devoicing process) but their obstruents are voiced in certain voiced contexts by the (later) application of the voicing process. For this pair of processes, application in the opposite order would be indistinguishable from suppression of the voicing process. But many pairs of processes are manifest in either order. For example, there is an assimilation process changing [t] to [s] before [s], and a process which I will refer to here as absorption, which changes vowel plus nasal to nasalized vowel before spirants. Both processes applied in Latin and Greek. Assimilation changed Latin [nepo:t:s] 'grandson' (gen. [nepo:t:s]) to [nepo:ss], which is simplified to [nepo:ss], and Greek [o:t-s] 'ear' (gen. [o:t-0s]) to [o:ss], whence [o:ss]. Absorption changed Latin [sanguin-s] 'blood' (gen. [sanguin-]) to [sanguis] (by subsequent denasalization) and Greek [hrin-s] 'nose' (gen. [hrin-os]) to [hri:s]. As the Latin form [mens] from [mont-s] 'mind' (gen. [mont-is]) shows, absorption could not apply after assimilation in Latin. Thus it could not apply to the [Vns] sequence which resulted from the assimilation of [Vnts], but only to original [Vns]. In Greek, however, the order of these processes was not limited at all: [himant-s] 'thong' (gen. [himant-os]) underwent assimilation and then absorption to yield [hima:s]. Since Greek did not limit the effect of either process, it can be assumed to reflect the innate relationship of these processes, whereas Latin reflects a limitation—by ordering—of the effect of the absorption process.1

I assume, then, that in its language-innocent state, the innate phonological system expresses the full system of restrictions of speech: a full set of phonological processes, unlimited and unordered. The most extreme processes are usually observable only in infancy: unstressed syllables are deleted, clusters and coarticulations are simplified, obstruents become lax stops, linguals become coronal vowels, and many pairs of processes are manifest in either order. For example,
which, although they are still nonsemantic, characteristically consist of well-articulated sequences of identical and stressed syllables composed of lax stop (or nasal) plus low vowel: [dadada], [pa-papa-pa-mama] or the like. Even this early there is some freedom, in that the stop may or may not be voiced by assimilation to the vowel, the vowel may or may not be fronted by a coronal, a coronal may or may not be palatalized, nasals may or may not be denasalized, and so forth. The first words resemble these post-babbling utterances in structure, and indeed they are often just continuations of these, with semantic import.  

Each new phonetic opposition the child learns to pronounce involves some revision of the innate phonological system. It appears that the mechanisms of this revision are the same as those which resolve contradictions between processes: suppression, limitation, and ordering. The child's task in acquiring adult pronunciation is to revise all aspects of the system which separate his pronunciation from the standard. If he succeeds fully, the resultant system must be equivalent to that of standard speakers.

In the view I am proposing, then, the mature system retains all those aspects of the innate system which the mastery of pronunciation has left intact. (But not only those: rules governing phonetically unmotivated alternations are certainly learned.) The processes which survive determine what phonetic representations are pronounceable in the language. For example, there is a process devoicing word-final obstruents (presumably a limited version of the process devoicing obstruents in general) which usually manifests itself immediately upon the acquisition of word-final obstruents. English-speaking children must suppress this process if their pronunciation is to conform to the standard, but German children need not, because German permits this devoicing: [hund/hunds] 'dog/dogs'. As the example shows, the devoicing process governs only the phonetic representation of German words, since the phonological representation of [hunt] is /hund/. In other languages it governs the phonological representations as well, in case there is no voicing opposition in morpheme-final obstruents. And in languages which, for example, lack morpheme-final consonants altogether, the process stays in the system but has no overt manifestation. This claim flies in the face of all phonological theories known to me, but it appears to be supported by the pronunciation, in such languages, of foreign words with final voiced obstruents, which, if they are pronounced at all, are characteristically devoiced.

Students of child language have noted striking regularities in the order in which phonetic representations are mastered. My studies have convinced me that these regularities can be fully explained by independently attested properties of the innate system—its processes, their inner hierarchies, and their interrelations—and by the three mechanisms whereby the innate system is revised. In particular, it appears that there is no need to refer to "implicational laws," such as Jakobson (1940) proposed, since to the extent that these are valid they seem to result entirely from the innate system.

Consider, for example, the implicational laws that affricates imply spirants and spirants imply stops (Jakobson, 51, 55). There seem to be only two general, context-free processes affecting these
obstruent articulations: obstruents become stops (which may be limited to affricates become stops) and affricates become spirants. If these processes are innate, it can readily be seen that there is no possible suppression, limitation, or ordering which could violate the implicational laws and thereby result in a different order of acquisition. In fact, even the "intermediate" levels of representation permitted by these mechanisms conform to the implicational laws. A similar account can be given for all the implicational laws of which I am aware.5

However, there are less general processes affricating or spirantizing palatal and even velar stops, and context-sensitive spirantizing processes affricating stops before high vowels, or spirantizing them after vowels, for example. As might be expected, if these apply there may arise contradictions to the order of acquisition predicted by the implicational laws. Jakobson was able to ignore these contradictions by interpreting the implicational laws in terms of phonemic representation, which could treat palatal affricates as stops, for example, if there was no contrast. But the contradictory context-sensitive processes cannot be ignored, for they may neutralize phonemic oppositions in certain contexts. Therefore the implicational laws cannot even account for the phonemic representation but only for the phonemic inventory, which is unaffected by contextual neutralizations.

Implicational laws, since if the processes are taken as the primitives of the theory, as I am proposing, it is possible to make predictions about representations as well as inventories, and in fact about representations at every level. Even if we extended the notion implicational law to allow for the contradictory processes, still the laws would themselves be contradictory, and to resolve the contradictions we would have to appeal to suppression, limitation, and ordering. In other words, there would then be no difference between the implicational laws and the innate processes, which is just what I am arguing.

Most modern students of child phonology have assumed that the child has a phonemic system of his own, distinct from that of his standard language. So far as I am aware, no evidence whatsoever has been advanced to support this assumption. There is, on the other hand, abundant evidence that the child's representations closely conform to adult speech. Since this claim is essential to the theory I am proposing—that the child's productions result from the application of the innate phonological system to some sort of phonological representation—it is appropriate to digress briefly from strictly phonetic concerns and examine some evidence of this phonological representation. Most major works on child language agree that the child has internalized a representation of adult speech which transcends in detail his own reduced productions. The most striking evidence, I think, comes when a child masters a phonological opposition he previously had merged. From that moment he pronounces the new segment in precisely the appropriate morphemes, without rehearing them, and the old substitute does not reappear again.6 Unless, of course, his mastery is only conditional, so that the process remains optional. But even in this case the variation between the new and
old segment will take place only in morphemes which in the adult lan-
guage have the former. Jakobson, discussing the emergence of [k] after
the suppression of the process that makes linguals coronal, says that
"when [k] finally appears, mistakes in the use of both phonemes ([k, t])
aris at first, especially those caused by a hypercorrect repression of
the expected [t] in favor of [k]" (54). But all the cases of this which
I have seen in the literature can be, in fact must be, explained as
the result of some additional process whose presence was not pre-
viously apparent. In the case in question, the culprit is the in-
fantile process which makes coronals velar if a velar occurs in the
word, so that 'cat' and 'dog', earlier pronounced [tat] and [tod],
change at the first appearance of velars to [kat] and [god]. Jakobson's
example of such a hypercorrection, quoted from Nadoleczny, is clearly
subject to this analysis: a child said *Duven Ta Herr Dotta, then Guten
Gag Herr Goka*, presumably for 'Guten Tag, Herr Doktor'. Far from being
evidence against the child's having mentally represented phonological
oppositions before he can produce them, such cases actually support the
claim. My son had a process deleting final stops, so that 'dog' was
pronounced like 'doll' at first, as [ds]. Immediately upon the
acquisition of velars 'dog' became [gd], the coronal being assimilated
to the deleted velar, while 'doll' remained unchanged.7

Returning to the acquisition of phonetic representation, I will
now cite some examples from children's speech illustrating the ordering,
limitation, and suppression of innate processes. Examples of
one child having ordered two processes which another child has not,
are common, but clear examples of a child actually performing the
ordering are difficult to find in the literature. However, I am
aware of a few cases. For example, Hildegard Leopold at 20 months
[dz] becomes [dz], and (b) obstruent voices before vowel (compare
[du] 'to, do'). But at 19 months 'choo-choo' had been [dud]
(Leopold 1947: 268 and 1939: 126), by unordered application of the
same processes. At 22 months Joan Velten devoiced word-final ob-
struents ([bat] 'bad, bite') and then denasalized everything after
a non-nasal consonant ([bub] 'broom', [bud] 'spoon'). But the only
word she had previously acquired which met the conditions of both
these processes, 'lamb', had been pronounced [bap], with the unor-
dered application making [m] into [b] and then [p] (Velten 1943).8
My son John had [kani] 'Channing' by (a) flapping of [n], flap-dele-
tion, and desyllabification. His pronunciation of 'candy' was at
first [kani], by processes (b) changing [nd] to [nn] and [nn] to [n],
applied before (a). Later 'candy' became [kani], by ordering (b)
after (a).9

Examples of limitation are much easier to find. Hildegard
Leopold voiced obstruents before voiced segments ([baba] 'papa')
and later only between voiced segments ([paba]), before she finally
suppressed the process ([papa], Leopold 1947: 31). By exactly paral-
lel steps, at 25 months Joan Velten first distinguished between
[pibi] 'paper, people, purple' and [pibi] 'baby', and then at 27
months between [papi] 'puppy' and [papi] 'probably' (Velten 290).
The denasalization process in her speech, mentioned above, was
limited to nonfinal positions before being suppressed: [sa:bud]
'salmon' becomes [sa:bun] and finally [sa:mun] (291). Chao's grand-
daughter, learning Mandarin, at first palatalized all coronals and
then limited this to continuant coronals (Chao 1951: 29f). 10

The [h]-deletion process, context-free in its most general form,
was successively limited in my son's speech to syllables of decreasing
stress, until he attained the standard formal pronunciation with [h]
dropped only before completely stressless syllables. This is a condi-
tional limitation even in the standard, at least in my dialect, and so
it may apply to increasingly stressed syllables in increasingly relaxed
speech (his henhouse, 'is henhouse, 'is hen'ouse, 'is 'en'ouse). The
example illustrates, incidentally, the way the inner hierarchy of a
process governs not only its form but its application; phonetic repre-
sentations requiring a violation of the hierarchy (e.g. 'enhou'ase) do not
occur.

As an example of suppression, I will cite just the word 'kitty'
as pronounced by a two-year old boy in successive interviews. His
pronunciation changed from [ki:] ([kili]) to [ki:] to [ki:] to [kiri] to
[kiti], by successive suppression of the processes of prevocalic tens-
ing, postsyllabic desyllabification, flap-deletion, and flapping (the
boy's parents speak a dialect with unflapped [t]). It is extraordinary
for a child to manifest all the intermediate steps between his first
and last pronunciations of a form. My son's pronunciation of this word
went directly from [ki:] to [kiri]. This does not necessarily indi-
cate that he suppressed all the processes at once, since suppression of
the first (flap-deletion) would have resulted in the same abrupt change.

Passy believed that phonetic change arises from "imperfect imita-
tion, by children, of the speech of adults" (1890: 225). This
hypothesis explains why, except in borrowing, phonetic change does not
seem to occur in adults. It also explains, as the Neogrammarian view
of phonetic change as due to subconscious drifts of pronunciation does
not, how change might be quite radical. And finally, to the extent
that children's deformations of adult speech are regular, it explains
why phonetic change is regular. If the child fails to master a certain
sound, it will appear that he has changed it to the sound he regularly
substituted for it. 11

This account can readily be expressed by the theory I am propos-
ing. A phonetic change occurs when the child fails to suppress some
innate process which does not apply in the standard language. Thus if
an American child fails to suppress the process devoicing word-final
obstruents, for example, his speech—compared to the adult standard—
will exhibit a phonetic change corresponding to the "addition" of a
devoicing process to the phonology. (This change has in fact occurred
in many dialects of English.) The change would not affect the phonological
representations of its originators, of course, nor would it affect
those of later generations if there remained any phonological support
for the original representations. In the dialects in question, 'bet'
and 'bed', earlier [bet] and [be:dr] by the standard process making
stressed vowels long or short before voiced or voiceless segments,
respectively, are still distinguished by vowel length as [bet] and
[be:t].
adjustment before devoicing. The innate, unordered application of these can be seen in the speech of children. For example, Joan Velten at first pronounced 'back' and 'bad' alike as [bat], with length adjustment applying after devoicing, then at 24 months distinguished these as [bat] and [bo:t], respectively (Velten, 289), by ordering length adjustment before devoicing. Thus her phonological system becomes identical, with regard to these processes, to the dialects cited above. Three weeks later she attains the standard pronunciation by suppressing the devoicing process.

The conservative influence of the standard exerts itself by rejecting most of the innovations of children. Innovations are only gradually admitted, often just conditionally at first. Thus it is that phonetic changes often begin as optional rather than obligatory pronunciations. The conservatism of the standard forces the innovator to suppress a process at least in his formal speech. Beside the dialects which have admitted obligatory devoicing, there are many others, more conservative, which still admit it only in relaxed speech. We must not imagine that the fact that phonetic changes are characteristically optional before they become obligatory means that they are at first optional in child speech. On the contrary, the child's progressions are essentially opposite to the tendencies of change—we might say the regressions—of the standard language.

This is why the typical progression from unordered to ordered application exemplified in the child appears in the opposite order in the corresponding phonetic change. Even those dialects which have admitted devoicing have required that it be ordered with respect to length adjustment. German has gone a step further in admitting unordered application, so that [bunt] has the same length whether it represents underlying /bunt/ or /bund/. The conservative Latin ordering of assimilation after absorption, as in [mens] ([mɛns]) was relaxed in popular Latin so that [n] was absorbed here as well.12

The child must limit a phonological process to the form compatible with the standard, and to the extent he fails, it will appear that he has "generalized" the process. The generalization will therefore conform to the hierarchies implicit in the process; change will proceed in the opposite direction along these hierarchies from the child's limitation of it. These hierarchies regularly assume the form of hierarchies of applicability if the generalizations are optional ones, as in the example of [h]-deletion cited above. They are sometimes even reflected in the isoglosses surrounding an innovating dialect. Ideally, the innermost dialects have admitted a process in its most general form, and each successive isogloss marks an additional limitation which has been enforced, until at the outermost isogloss we encounter dialects which have not admitted the process in any form. Of course, the ideal is rarely encountered. The classic example involves the isoglosses of the "Rhish Fan," on the boundary between High and Low German.13

I have dwelt on phonetic change at some length because the account of phonetic acquisition presented here appears to explain fully all the currently known mechanisms of regular phonetic change. The apparent addition, generalization, and unordering of processes arise in the child's failure, respectively, to suppress, limit, or order processes of the innate system to the extent required by the standard language.
In other words, the child simply fails to master a phonetic opposition in some or all contexts. This is really all that is involved. But we must refer to the innate system and the mechanisms by which it is limited to understand the precise nature and the regularity of phonetic change.

The theory of the acquisition and change of phonetic representations outlined here can be extended in a natural way to account for the acquisition and change of phonological representations (Stampe 1968). And finally, it can account for implicational regularities such as Jakobson observed in the phonological systems of the world's languages, but in much finer detail than can Jakobson's theory, which was limited just to a subset of context-free phonological processes. The markedness theory of Chomsky and Halle (1968: chapter 9) was limited to approximately the same set of processes. As a result both these theories were limited just to the most underlying levels of representation, which are relatively, though not entirely, undisturbed by the remaining processes. An important difference between those theories and the theory outlined here, then, is that the latter is intended to account for all levels of phonological representation. Of course, such theories can succeed only insofar as they are made substantive, in the present instance by a description of the innate phonological system. It becomes increasingly obvious that this staggering task must confront any serious attempt to advance our understanding of phonology.

The Greek processes are unordered in that the correct forms would result if the processes applied sequentially and iteratively in random order. The notion "unordered" corresponds to the "unmarked ordering" of Kiparsky (1965) except that I extend it, as noted, to include re-application. (More on this below.) "Ordered" here corresponds to Kiparsky's "marked ordering."

However, most of the processes mentioned in this paragraph make limited appearances, at least, in some adult languages. For example, Vietnamese has lost Austroasiatic unstressed syllables, Oceanic has lost Austronesian consonant clusters, etc. Such radical changes result from the collective effects of numerous specific processes.

"Post-babbling" utterances, as defined here, are to be distinguished from the unstructured, random vocalizations of true babbling, which, significantly, is essentially alike in deaf and hearing children. Since the structure of post-babbling utterances can be accounted for by the innate phonological system, one might further speculate that they are underlain by phonological representations, in some sense, perhaps as crude imitations of adult speech, prior to the recognition of its distinctions and semanticity.

This crucial issue is discussed in Stampe (forthcoming), where it is contended that the problem of "phonological admissibility" is contained in, and therefore inseparable from, the larger problem of loan phonology.

Conversely, it turns out that incorrect implicational laws, like Jakobson's conjecture that spirants imply nasals (Velten 1943: 282), fail to correspond to actual processes; in this instance, there is no general process changing (non-nasal) spirants to nasals.

The most notable exceptions to this generalization involve words which the parents have imitated. When adults adopted my son's [pibi] for 'T.V.' (by despirantization and labial "harmony"), this word persisted for months after he had suppressed these processes, even though he could pronounce [tivi] with ease. Later after he adopted the standard form he still occasionally used [pibi] in babbling. The most striking cases of "frozen speech" involve elder siblings, who occasionally adopt some stage of the speech of the younger, and continue it as a private language even after the younger sibling has attained standard pronunciation. (Jakobson 1940 cites some references.) Such exceptions support, rather than refute, the claim that the child's representations correspond to the productions of his models.

In "Yes, Virginia..." I argue that the child's phonological representations must in fact be at least as deep as a "phonemic representation" of adult speech, based on examples like [badn] for adult
discussed in StamPa forthcoming.

KiParskY (1968), that a process is not simply reordered but rather added in the first place. The more recent proposal by Halle 1962 that change reflects the simplicity of these rules so that (other things equal) the fewer things being constructed by the child to account for his linguistic experience, brings no support to the idea that the girl's phonological representations were a phonemic version of her own pronunciations.

This view was applied by Jakobson (1940) to the phonemic level. The more recent proposal by Halle (1962), that change reflects the addition of a process to the phonology, is not so limited, and reflects our growing conviction that phonological processes are not mere descriptive devices but rather genuine components of the mental grammar. But Halle's implication that adults might spontaneously add a process is difficult to understand. Halle's general theory is based on the assumption that all phonological processes are rules which are constructed by the child to account for his linguistic experience, and that the phonological system is evaluated according to the simplicity of these rules—so that (other things being equal) the fewer rules, the better. It is not at all clear, given this view, why a process should be added in the first place.

On the basis of this Latin example it can be argued, contra Kiparsky (1968), that a process is not simply reordered, but rather unordered in this sort of change, because the absorption process had to apply twice. The first, and original, application ([sanguis < sanguin-s]) left no nasality behind, due to application of an unordered process denasalizing vowels. But the second application did leave nasality: [mûs < mens < ment-s]. That is, when absorption was unordered so that it could apply to the result of assimilation, the denasalization process became ordered. (The ordering is an incidental result of change, not a primary change. Similarly, an unordered process devoicing final obstruents might become ordered upon the addition of a process deleting final vowels. Naturally, further changes might undo these conservativeness.)

This suggests that Kiparsky's explanation of order changes as simplifications of the grammar is incorrect (compare also note 14).

It is possible to avoid this result only by denying the relationship of the two absorptions; but it was to express such relationships that Kiparsky proposed the notion of change by reordering in the first place. For further examples of this sort, including "iterative" application, see StamPa (forthcoming).

It should be emphasized, perhaps, that "generalization" as used in this discussion has little relation to its use in generative phonology theory, since the latter is concerned not with innate processes but rather with rules which are "internalized" by the child to represent "significant generalizations" about his language. That the child supplements the innate phonological system with rules, particularly morphological and expressive ones, is not at issue. But so little is known of such rules as yet that most recent speculations about constraints on them—such as the notational and "markedness" conventions of Chomsky and Halle (1968: chapters 8 and 9)—seem quite premature.

Kiparsky (1968) also proposes change by suppressing processes of the standard language, which he explains, along with generalization and reordering, as simplification of the grammar. He is thus unable to explain the addition of processes, whereas my proposal can explain addition, generalization, and ordering but not suppression. But there are less than a half dozen clear cases of change by suppression, against innumerable cases of addition.

One of the clearest cases involves the suppression of the German process devoicing word-final obstruents in certain Yiddish and northern Swedish dialects (Kiparsky 1968: 172). That the process was suppressed seems certain, but the cause appears not to have been "simplification" but rather a dilemma occasioned by the loss, in all these dialects, of most word-final schwas. This introduced thousands of voiced obstruents into word-final position and, since the lost schwas lacked clear morphological support, flatly contradicted the devoicing process. Two resolutions were possible to children confronting this situation. They could apply devoicing to the newly final obstruents, and thus merge [bunt] and [bund < bundo] as [bunt]; or suppress it, and keep [bunt] and [bund] distinct. Some dialects took the former option and others, the ones Kiparsky cites, the latter. This suggests that although suppression occurs under certain circumstances, it is not a primary mechanism of change.

Of course, a process may be "suppressed" if a later process is added whose effects include its own. (This situation is exactly parallel to the unordering of a process, if there are no intervening processes—like the denasalization process in note 12—to reveal an earlier application.) But in the theory I am proposing, the first process would not really be suppressed—since it would not interfere with the pronunciation of the language in any way—but would merely have become unobservable.

The claim that change by suppression is extraordinary is equivalent to the claim that the child does not ordinarily suppress processes which apply obligatorily in the standard language.
Although this superficially appears to occur in child language, a deeper study seems invariably to reveal that such appearances are due to other factors, most typically a distinct process which, applying later, undoes the effects of the standard process.

REFERENCES


Stampe, David. Forthcoming. *A Dissertation on Natural Phonology."

A DISSERTATION ON NATURAL PHONOLOGY
A phonological process is a mental operation that applies in speech to substitute, for a class of sounds or sound sequences presenting a specific common difficulty to the speech capacity of the individual, an alternative class identical but lacking the difficult property.

For example, voiced stops are relatively difficult to articulate because their characteristic obstruction of the nose and mouth impedes the air stream on which the glottal vibration of voicing depends. There is a phonological process which avoids this difficulty simply by substituting voiceless stops for voiced; it is observable in the speech of many young children and in the pronunciation of voiced stops by speakers of languages which lack them—Hawaiian, for example. There are other processes which offer means of avoiding the difficulty without giving up voice: these include (pre)nasalization, which relieves the supraglottal pressure through the nose; spirantization, which relieves it through the mouth; and implosion, which relieves it by lowering the glottis and increasing the size of the supraglottal cavity.

An example involving a class of sound sequences: a sequence of nasal plus spirant, e.g. [/ns], is difficult to articulate because it requires the release of the oral closure of the nasal to coincide precisely with the closure of the velum. If the velum closes
before the oral release, there is in the interim an oral stop articulation, resembling [nts] or [nds]. If the oral release precedes the velar closure, there is in the interim a nasal spirant articulation, resembling [nzs] or [n̩s]. There are two processes responding to this difficulty of timing. One inserts an oral stop: [nts]. The other substitutes for the nasal stop a nasal lacking oral closure: [โนะ]. Either process is observable in children’s speech and in the pronunciation in various languages of nasal plus spirant sequences, e.g. in the variation [dents, d̩ns] dance of some English speakers.

The inborn capacity for speech seems to be uniform among peoples of all genetic heritages: normal individuals of whatever birth seem to be capable of acquiring the pronunciation of any language they confront as children. On the negative side, this inborn capacity seems to reveal similar limitations the world over. All children’s earliest speech typically consists of sequences of simple consonant-vowel syllables: the consonant and vowel, lacking any positive articulatory properties, are respectively a nonnasal voiceless unaspirated (etc.) stop and a stressed variety, such as [a̩], of the neutral vowel. These limitations are not absolute: children overcome some of them in their earliest utterances, and within a few years normal children have mastered most of the sounds and sound sequences of their language.

But in the meantime children are not reduced to silence. Confronted with sounds they cannot pronounce, they employ substitutes. These substitutes are not merely random or occasional, but are regular and unexceptional in the child’s speech. One child who cannot pronounce lamb says [zab], another says [j̩æ]. But the child who says [zab]—Joan (Velten 1943)—pronounces all initial [l]’s as [l̩] (e.g. [l̩a] leaf, [lu̩l̩] belong), all nasalized [æ̩]’s as nonnasal [a] (e.g. [æ̩ts] for [æ̩ts] ants, [han] for [hænd] hand), and all nasals following nonnasal consonants in the same word as voiced stops (e.g. [sæ̩b̩] salmon, [zæd̩] lion). The child who says [j̩æ̩]—Daniel at fifteen months (from my notes)—pronounces all initial [l]’s as [l] (e.g. [j̩æ̩t̩] light, [j̩i̩v] leaf) and all sequences of vowel plus nasal as long nasalized vowel (e.g. [sæ̩k̩] some, [mæ̩] man).

This regularity might be explained as a direct result of physical limitations, overcome eventually by maturation. However, the substitutions themselves are not merely peripheral physical phenomena. It is not uncommon for children to produce a sound correctly at first and later submit it to substitution: Hildegard’s [priti] pretty (Leopold 1939:120-1) is an interesting example; the expected pronunciation [biti] (compare [baba] papa) did not appear for almost a year. Children who substitute [w] for [r] in speech may use [r] in play or to imitate a growling dog, as Jakobson (1968:25) notes. Alexandrov and Sully (Jakobson 1968:15) observed children who dropped [j] but substituted [j] for [l]. Joan (Velten 1943) devoiced final stops, as in [but] bread, bed, bird, but used voiced stops finally as substitutes for nasals: [bud] spoon, bin.

Finally, the child’s repertoire of sounds may remain unchanged, yet he may change his substitutions: Joan at 15–33 months could
pronounce only the vowels [a] and [u], but up to 22 months for
[e] she substituted [a], and thereafter [u]. My son John used
[j] for initial [l] until four years, and thereafter [w], though
he was equally able to say [j] and [w] throughout this period.

The conclusion that these substitutions are mental can also
be drawn from the fact that they occur regardless of how or whether
sounds are actually physically pronounced. Children often pro-
nounce vowels long before voiced segments, as in standard English
[be-d] bed beside [bet] bet. It is not unusual for children to
lengthen vowels even before segments that they devoice, e.g. Joan's
[bu-t] bead versus [but] beat (Valten 1943:289) or A's [la-i-t] slide
versus [lait] light (Smith 1970:42). After voicing is acquired,
the same words have lengthening even when they are whispered, even
though this makes the causal segments physically voiceless. In
fact, a bit of introspective phonetic reveals that bed and bet
retain their length distinction even when they are pronounced
silently in our thought. There is therefore no doubt that these
substitutions are mental. We must look elsewhere for an explanation
of their regularity.

It seems to me that it would be mistaken to seek a phonetic ex-
planation, much less a physical one, for the regularity of substi-
tution, whether in children or adults. One has only to consider
the chaotically unpredictable pronunciations of a specific single
word, e.g. lamb, which would result if the child avoided [1] by
using [j] in one utterance, [w] in another, [n] in another, and
so forth—to cite only a few of the substitutions observable in
various children. This is not to deny that a great deal of vari-
ability is tolerated even in adult pronunciation: there are a
dozen or so pronunciations of a word like divinity (Chapter II)
and over three dozen pronunciations of the modal can in my speech,
for example. But unlike the variants of the word lamb envisioned
here, these actual variants respond to various speech tempos and
various contexts in a very regular way.

Chaotically variable substitution occurs, if at all, only in
earliest childhood. Compare, for example, the various treatments
of [s] by Hildegard in her first year of speech (Leopold 1939,
under s in the word-indices). According to investigators I have
discussed these matters with, there are two situations that en-
courage massive variability. One is the habitual insistent cor-
rection of the child's pronunciation by adults, which seems f o
interfere with the child's tendency to habituate particular sub-
stitutions. The other is a lack of opportunity for the child to
have conversation, which deprives him of the feedback he needs to
determine whether his speech is understandable or not.

I have argued that phonological substitution is not merely
a reflection of peripheral motor constraints. The examples given
above with [nts] or [Vs] substituted for [ns] seem to invite a
motor interpretation, since either might result directly from a
physical mistiming of the articulation of [ns]. But typically one
or the other substitute is regularized in a language. Even in
English dialects where they occur in variation, speakers have the
option of using the difference to distinguish prince from prince,
for example, as [priŋs] versus [prints]. In folk spelling the substitutes are likely to be written even where this goes against the underlying phonological structure of the language. In teaching Soras to write their language (one of the Munda languages of India) I found that they usually wrote [nts] for what was clearly underlying /ns/ in the language, as in [kan] 'animal classifier' plus [sim] 'chicken', pronounced [kantsim], although triple consonant clusters do not otherwise occur in the language. Children's spontaneous spellings of English dialects which lose the oral closure of the nasal before spirants and voiceless segments omit the nasal (Read 1971:18-19), indicating an awareness of its absence in speech. Such evidence indicates clearly that where [nts] or [ʌs] appear for [ns] they are not merely motor slips for an intended [ns], but represent distinct phonetic targets supplied by mental substitution. This is confirmed by the difficulty, for speakers whose idioms require these substitutions, of pronouncing [ns] even in silent mental speech.

Although phonological substitution is a mental operation, it is clearly motivated by the physical character of speech—its neurophysiological, morphological, mechanical, temporal, and acoustic properties. This is now becoming widely recognized, and I will confine my argument here to an outline of some of the major types of evidence. First is an observation made explicit chiefly by Morris Halle, that substitutions refer to classes of sounds definable in terms of common articulatory-acoustic properties, or features. The class of sounds undergoing substitution, the class substituted, and the class of sounds—if any—in whose neighborhood the substitution occurs are all definable in terms of these features. Occasional doubts have been raised about the truth of this, but most of them have been resolved by independently motivated revisions of our ideas of which phonological features there are and what they are like.

A second sort of evidence for the physical motivation of substitutions comes from substitutions dependent on neighboring sound-classes. Typically there is a physically definable connection between such a substitute and its context. Traditionally recognized relationships include assimilation, where the substitute takes on a feature of the context, and dissimilation, where the substitute loses a feature of the context. In some context-dependent substitutions the feature changed is not identical to the contextual features, but there is nonetheless a physical interdependency between them—for example, between stress and aspiration in the substitution (in English, for example) of aspirated for nonaspirated voiceless stops in the onset of stressed syllables.

A third type of evidence comes from so-called optional substitutions, whose application typically depends on how much attention is given to articulation. In relaxed speech in any language there are many substitutions which do not occur in formal styles. These substitutions often apply to words and phrases which are readily recognizable due to their grammatical or semantic status or their commonness in conversation. In my speech the verb think is optionally
pronounced with [h] instead of [θ], but only in its commonplace parenthetical use, as in I think it's raining, isn't it?—not in its use as a main verb, as in I think, therefore I am. The common phrase I don't know can be reduced to [θəʊ][θ] or less, but the phonologically similar but uncommon phrase I don't guess tolerates no such reduction. Although such examples establish the function of phonological substitutions in adapting speech to casual articulation, the limits placed on substitution by lexical and statistical considerations remind us that the application of phonological substitutions is nonetheless mental.

Fourth, phonological substitutions exhibit degrees of generality according to the degree of physical difficulty involved in the articulation of the various sounds to which they potentially apply. In the assimilative substitution of palatal for nonpalatal consonants adjacent to palatal vowels, a substitution is more likely next to vowels which are extremely palatal (e.g. [i]) versus [e] and which are in close contact with the consonant (e.g. not separated by a syllable boundary), and it is more likely to affect more readily palatalizable consonants (e.g. [s] versus [ʃ]). Similarly, the loss of palatality in vowels which occurs, for example, in vowel neutralization, basically a context-free process, is more likely to affect vowels of weaker palatality (e.g. [ə], versus [i]) and weaker articulation (lax, versus tense; unstressed, versus stressed).

These and other widely recognized properties of phonological substitutions, including the fact that they are far more extensive in the speech of children, leave little doubt that they are ultimately motivated by the physical character of speech. Although substitutions are mental in occurrence, they are physical in teleology: their purpose is to maximize the perceptual characteristics of speech and to minimize its articulatory difficulties. Phonological processes are mental operations performed on behalf of the physical systems involved in speech perception and production.

One puzzle this leaves is why mental speech which is not actualized physically should be subject to phonological substitutions, if these are physically motivated. This is a complex topic, involving not only speech but also other ordinarily physical activities which can be performed in the mind by those who have mastered the corresponding physical activity. In my opinion, mental speech (as distinct from certain other sorts of thinking which employ language) is simply a sublimated variety of physical speech. Its tempo is not strikingly faster than that of physical speech, and its rhythm is essentially identical. The sublimation process seems to be difficult or impossible for young children, who typically accompany their play with physical speech, which later becomes reduced in volume, and eventually shows no manifestation at all outwardly, as they grow older. Similarly, reading is done aloud at first and only later becomes silent. I cannot pursue this interesting topic further here, but since it seems that mental speech may be physical in origin, there seems to be no difficulty in attributing the phonological substitution that it shares with
physical speech to its physical origins.

Sounds and sound-sequences are not all equally easy to articulate. The more demanding ones are relatively rare in languages of the world and are usually acquired late in childhood. This, and the converse of it, that less demanding sounds are relatively common and are acquired early, has been documented by Jakobson (1968). But when we consider the incredible diversity of sound patterns in the world's languages, we must conclude that aside from some sounds which are almost universal and others which are extremely rare, most sounds and sound-patterns must be of relatively equal difficulty. There are languages with elaborate vowel systems and relatively few consonants (e.g. in the Mon-Khmer group), languages with rudimentary vowel systems and scores of consonants (e.g. the northwest Caucasian languages), languages with few sounds but many sound-combinations (various northwest American Indian languages), languages with many sounds but few combinations (e.g. some South American languages).

The retention of tonal distinctions in some languages of east Asia, South America, and Africa with impoverished syllable inventories seems to indicate that there are limits to the simplicity of sound-patterns and inventories, beyond which homophony becomes intolerable. Another indication of this is the systematic development of extensive compounding, as in Chinese and the South Munda languages, which seems to originate in the attempt to disambiguate homophones. Conversely, it is probably not accidental that we find no languages combining the large vowel system of, say, Nicobarese with the huge consonant system of Ubykh, and permitting the complex combinations of sounds of, for example, English. Broadly speaking, languages seem to require, and to settle for, a certain moderate phonological capacity.

The diversity of systems with which this capacity is achieved is testimony to the great flexibility of the speech faculty of human beings. The same flexibility can be seen even in children. Despite the opinion of early investigators, it is now generally recognized that different children acquire sounds and sound-patterns in quite different orders. This is not to deny that certain orders of acquisition are fixed. But there is much more diversity in children's sound systems and their development than was at first recognized.

However, much of this diversity is only apparent. Detailed analysis of the distinct substitutions of distinct children often reveals that these are due to a small number of phonological processes shared by many of the children but applying in some of them in a cumulative manner. Consider Joan and Daniel's pronunciations of lamb as [za]b and [ji`,] respectively. Joan's surprising [z] for [l] turns out to result from three distinct processes, two of which are shared by Daniel and all of which can be paralleled in other children and in languages of the world.

Both children delateralized [l] to [j]. This process, which is also responsible for the pronunciation of dark [t] as [w] in both children, has been studied in detail by Edwards (1970). She

Whereas Daniel used the [j]-substitute in his actual pronunciation of [l], Joan spirantized [j] to [ç]. This process does not apply in the speech of Daniel or the other children cited above, though at an earlier stage Hildegard did apply it briefly, e.g. [jou] you (Leopold 1939:34). It is exemplified diachronically in many languages, e.g. in French [ç] as in Jules for Latin [j] as in Julius, and is an intermediate step in the further change of [j] to [ç], as in Italian Giulio.

Finally, Joan depalatalized alveopalatals to plain alveolars, changing [ç] to [ç]. Velten states that she pronounced adult [ç] as [ç], but cites no examples. My analysis of the substitution of [ç] for [l] entails, in its three steps [l] + [j] + [ç] + [ç], that she should have pronounced adult [j] as [ç]. This is confirmed by yard (Velten 1943:290). The final depalatalization process is paralleled in Joan's [was] for [waf] wash, [mats] for [mat] mitch, [tabudz] for [kabidz] cabbage and [garbidz] garbage. This process is shared by Daniel, e.g. [bas] brush, [ṭits] teacher (Edwards 1970); it does not affect [l] as in Joan's speech because he does not spirantize the [j] substituted for [l]. Depalatalization occurs also in A's [du] shoes, with the depalatalized [s] stopped as in [du:n] soon (Smith 1970), my son John's [su:z] shoes and [dzaz:n] John, Melissa's [bwaz] brush (Edwards 1970), Edmond and Charles' [sâbr] chambre and [bezer] bergère (Grégoire 1937:345), substitutions of [s, ç] for [ç, ç] observed in French and Czech aphasics (Jakobson 1968:61), and the change of the palatals [ç, tç, dç] to [s, ts, dz] in the Desia dialect of Oriya (observed personally in the Ganjam and Koraput districts of Orissa, India).

I cite these parallels, whose number could easily be multiplied, to show that the individual processes delateralization, spirantization, and depalatalization are matched in a variety of child and adult systems. On the other hand there seems not to be a single case except for Joan in which [l] becomes [ç], or even [ç]. The only explanation for these facts is the one I am proposing: that Joan's pronunciation of [l] as [ç] is due not to a process peculiar to her own speech—apparently no such process exists—but rather to a sequence of processes common to many children and languages.

Furthermore, only on this assumption is it possible to understand the phonetic utility of Joan's pronouncing [l] as [ç], for there are scores of languages which have [l] but lack [ç] (Hockett 1955). The function of the individual processes responsible for her unusual substitution, on the other hand, seems fairly clear. The function of delateralization is rather clearly to eliminate a tongue-configuration whose difficulty makes laterals some of the last sounds children master (Jakobson 1968:57); this configuration is avoided not only by this process, which substitutes glides, but
also by processes which substitute \([d], [n], or [r]\) for \([l]\) (Edwards 1970). Spirantization increases the audibility of \([j]\) by substituting a sibilant \([\jmath]\). Its function becomes apparent when we note that the most common alternative substitution of \([j]\) is null, as in Jennifer's \([u\cdot j\, y\, o\, u\, n]\), Elena's \([u\, k\, l\, o\, o\, k]\) (Edwards 1970: 23), and the Scandinavian loss of Germanic \(*[j]\), as in Icelandic \(a\, f\, j\, a\)'year, Jahr'. Finally, dePafatalization eliminates the fronted and raised tongue-posture of \([\jmath, 3, etc.]\) in favor of the neutral posture of the plain alveolars \([s, z, etc.]\).

These imprecise remarks are only intended to be suggestive of the ultimate phonetic explanations of the individual processes in question. Despite some promising beginnings in the nineteenth century, phonetics has unfortunately not concerned itself with the explanation of phonological processes. But I think that this neglect is understandable so long as phonology fails to analyze gross and inexplicable substitutions like \([l]\) for \([l]\) into their constituent processes. The correct analysis of such substitutions requires, except in the most serendipitous cases, a careful comparison of parallels in a variety of child and adult phonological systems.

The failure of previous phonology to carry through such analysis is due in part to the structuralist conviction that languages are to be understood in terms of their own structure. In phonology this led to a rejection of the traditional view that phonemes are underlying mental sounds in favor of the functional view that they are abstractions based on the distinctive features of actual sounds. In this view phonemes are related to sounds not by the application of processes but by the invariance of their distinctive features. However, there are well-known examples of phonemes whose phonetic realizations 'overlap', e.g. the \(t\) of Danish, pronounced \([d]\) in posttonic positions, exactly like pretonic \(d\). (Jakobson's analysis of \([t]\) and posttonic \([d]\) as 'strong' versus 'weak' pretonic \([d]\) and its posttonic congener \([\delta]\) (1962:424) appeals to an otherwise unjustified distinctive feature.) The posttonic pronunciation of Danish \(t\) and \(d\) as \([d]\) and \([\delta]\), respectively, is due to voicing and spirantization processes.

Furthermore, the same processes apply in Tamil, for example, where \(t\) is both voiced and spirantized to \([\delta]\) in postvocalic position (Trubetzkoy 1969:141). In Danish the effects of these processes are not cumulative: underlying \([d]\) is spirantized to \([\delta]\), but not the \([d]\) which results from the voicing of \([t]\). In contemporary phonology this is explained as due to the ordering of spirantization in Danish prior to voicing, whereas in Tamil the application of spirantization is unrestricted with regard to voicing.

What has not been adequately recognized in discussions of ordering is that its recognition depends on our assumption that certain phonological processes are possible and others impossible. In the Danish example the facts could be described by positing a single process which simultaneously voices postvocalic stops and, if they are already voiced, spirantizes them. In fact such an analysis would be simpler from the point of view of Danish in that the postvocalic context of both changes would be expressed just once. But from a language-universal point of view, this analysis
is not satisfactory: it merely adds another process to the many we have to explain the existence of. For there are languages in which voicing and spirantization occur separately: in American English [t] is voiced intervocally, but [d] is not spirantized; in Old Irish [t] and [d] were spirantized postvocally, but [t] was not voiced. The analysis of the Danish situation as resulting from precisely the same processes as are found in American English, Old Irish, Tamil, and many other languages, but with an ordering imposed in Danish which is not imposed in Tamil, brings together in just two processes a diversity of phonological systems. And it eliminates a putative process which simultaneously but selectively performs two quite distinct phonetic modifications.

Furthermore, this language-universal analysis provides us with an interesting prediction: that in no phonological system will [t] become [ð] unless [d] becomes [ð]. This follows from the assumption that the change of [t] to [ð], as in Tamil, involves two distinct processes, voicing and spirantization. If this is so, there is no way that [t] could be voiced and spirantized without [d] also being spirantized. This prediction is not overturned by any language of which I am aware.²

The same reasoning reveals that processes may be ordered in children's speech. My analysis of Joan's pronunciation of [l] appealed to two processes, delateralization of [l] to [j] and spirantization of [j] to [ʃ], and, as the examples cited show, Hildegard shares both of them: [ʃu] you, [juʃ] lustacht. But for Hildegard [l] remains [j], it does not become spirantized as in Joan's speech. The difference between their pronunciations lies not in the processes but in their order of application. Hildegard does not apply spirantization to the output of delateralization; Joan does. Thus whereas Joan cannot pronounce [j] at all, Hildegard can pronounce it if she attempts to say [l]. This apparent oddity is no stranger than the Dane's inability to say postvocalic [d] except by aiming at [t]. Not surprisingly, Hildegard soon acquired [j] per se; from withholding the application of spirantization to [j] derived from [l] it is only a short step to withholding its application to any [j]. But her substitution of [j] for [l] persisted, revealing that, as our analysis implies, her earlier treatment of [l] and [j] was not a single complex process but two distinct simple processes.

If one were the god of language, free to create any sort of system for mortals to adapt difficult sounds to their limited speech capacities, one would surely reject a system involving such atomistic responses to specific difficulties as these individual processes represent, and instead attempt to devise a unified, global system. However, nature has not managed to furnish such a system. The pronunciations of children and of adult languages reveal themselves to be the result of numbers of substitutions which, however phonetically natural they may seem individually, have collective results which are perversely cross-purposeful.

For example, there is a context-free process which denasalizes vowels, well-attested in children (Jakobson 1968:71-72) and in languages of the world, for which various plausible phonetic explanations have been proposed. There is in addition a process,
also widespread in children's and adult's speech, which nasalizes vowels adjacent to nasals, and which likewise is capable of phonetic explanation, in this case involving the timing of sequences of articulatory gestures (Schourup 1972). Given the distinct teleologies of these processes—one eliminating a difficult feature (nasality) in vowels in general, the other permitting adjacent nasality to spread into a vowel—it is not too surprising that they make opposite substitutions. What is surprising is that in most languages both these processes apply, first denasalization and then nasalization. American English is an example. Foreign words with nasalized vowels are typically denasalized by English speakers: French [manɑ̃] becomes [man], etc. At the same time, vowels before nasals are nasalized not only in native alternations like [si] beside [si•see] but even in foreign words with oral vowels before nasals, e.g. [bonɑ̃] bon am', which in the English mouth becomes [bɔnɑ̃] or the like.

It is tempting to suppose that these restrictions on nasality are somehow unified into a single consistent restriction: that vowels become nasal before nasals and nonnasal before nonnasals. However, there is no evidence of such a unified rule in any language known to me. Furthermore, the assumption that vowels in general are denasalized in English enables us to account for the fact that English speakers identify their vowels as nonnasal. The superficial nasality of vowels before nasals is not perceived as such, even in words where only this nasality remains to indicate the underlying nasal, e.g. [kɔt] can't versus [kaT] cat. Vowel denasalization is in effect a condition on underlying representation in English (and in most languages), with vowel nasalization affecting only surface realizations of vowels. It has been customary to view constraints on underlying representation as due not to processes—potential substitutions—but rather to 'redundancy conditions', 'markedness conventions', or the like—essentially negative conditions which prohibit certain representations but do not specify a surrogate when they are confronted in foreign words. I think the fact that speakers of English denasalize vowels which cannot be explained as derivatively nasalized, as in the French examples cited above, dictates against such a distinction. Speakers do not substitute just any similar pronounceable sound, e.g. syllabic [ŋ], for a nasalized vowel, and this fact is adequately explained by the fact that there is no general process making such a substitution, whereas there is a general process denasalizing vowels. Thus Joan Velten, for example, says [na] for French [nɑ̃] non, [ats] for [ɑ̃ts] ants, [han] for [nɑ̃d] hand.

As further testimony to the individualistic purposes of phonological processes, it can be shown that prior to vowel denasalization, Joan applied vowel nasalization. There are two pieces of evidence for this, both involving intervening processes. One comes from forms like [nan] lion, in which the [l] is probably nasalized by a nasalized vowel—[æ][n] + [lən] + [ʔən] + [nən], with vowel denasalization to [nan]—since there is no evidence known to me that this sort of nasalization can be caused by non-adjacent nasality.
The other evidence involves a context-dependent denasalization process which denasalizes all segments following a non-nasal segment in a word: [sabud] salmon. This process clearly follows [l]-nasalization, otherwise we would have [zad] for lion. (In fact when Joan stops nasalizing [l], this pronunciation replaces [nan].) The peculiar thing about the progressive denasalization process is that it is not triggered by vowels, e.g. [am] N. We can account for this by assuming that the vowel nasalization process implied in [nan] + [nān] + [nīn] + [īn] lion applies prior to progressive denasalization, since [ān] + [am] N would then contain no nonnasal at the stage of derivation at which progressive denasalization applies, and that vowel denasalization (as in [na] non, etc.) applies after progressive denasalization. This assumption allows us to account naturally for a peculiar condition on progressive denasalization which Velten noted in Joan's speech (286), that although it was triggered by nonsyllables [g, d, z, w, etc.], [h] caused denasalization only before [u], not before Joan's other vowel, [a]: [hub] home, like [wub] room, beside [ham] ham. Now [h] is, phonetically, a breathy nonsyllabic copy of the following vowel; [h] in Joan's pronunciation of home is presumably [ūh] whereas [h] in ham is [āh]. It is well established that the nasalizability of vocalics is a function of their syllability (Schourup 1972) and an inverse function of their height (Chen 1971). Accordingly, Joan's vowel nasalization process seems to be limited to syllables and non-high nonsyllables, namely [āh]; it does not extend to high nonsyllables such as [w] or [ūh]. Consequently, while the nonsyllabic of ham [gāhm] is nasalized to [gāhm], that of home [gūhm] and room [wūhm] is not, and it causes progressive denasalization: [wūh], [wub].

To summarize, Joan has (1) iterative regressive nasalization of syllabic or non-high sonorants before nasalized segments (this would include [l] as a subcase), (2) nasalized [r] becomes [n], (3) iterative progressive denasalization of all segments after a nonnasal in a word, and (4) denasalization of vocalics:

<table>
<thead>
<tr>
<th></th>
<th>room</th>
<th>salmon</th>
<th>home</th>
<th>ham</th>
<th>lion</th>
<th>non</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x)</td>
<td>wum</td>
<td>samun</td>
<td>hum</td>
<td>ham</td>
<td>ian</td>
<td>nā</td>
</tr>
<tr>
<td>(1)</td>
<td>wūm</td>
<td>sāmūn</td>
<td>hūm</td>
<td>īn</td>
<td>nām</td>
<td>nā</td>
</tr>
<tr>
<td>(2)</td>
<td>wūm</td>
<td>sāmūn</td>
<td>hūm</td>
<td>īn</td>
<td>nām</td>
<td>nā</td>
</tr>
<tr>
<td>(3)</td>
<td>wub</td>
<td>sabud</td>
<td>hub</td>
<td>īn</td>
<td>nām</td>
<td>nā</td>
</tr>
<tr>
<td>(4)</td>
<td>wub</td>
<td>sabud</td>
<td>hub</td>
<td>īn</td>
<td>nām</td>
<td>nā</td>
</tr>
</tbody>
</table>

Of these processes each can be paralleled in other children and languages, and two are part of English (vowel denasalization and iterative regressive nasalization), though in a reverse order from that in Joan's speech. She corrects the order a year later (Velten 291), having suppressed the un-English nasalization of [ī] (2) and progressive denasalization (3) in the meantime (287n, 291).

The existence and independence of contrary processes like vowel denasalization and vowel nasalization must be sought ultimately in their contrary phonetic teleologies. Context-free processes—like vowel denasalization, which eliminates a feature that interferes with the articulation and acoustics of vowels—respond to the inner complexities of single segments. Segments are often
compounded of properties which compromise each other, and therefore there are processes which eliminate these compromises, emphasizing one property at the expense of another. For example, the vowel [a] may be viewed, with the ancient Indian phoneticians, as a compound of palatality or [j]-ness with sonority or [e]-ness. Since palatality is maximized by narrowness, and sonority by openness, the vowel clearly represents a compromise of these properties. There is one process which increases its palatality at the expense of its sonority: (1) Raising, [e] → [i], narrows the vowel. Two processes increase its sonority at the expense of its palatalitY: (2) Lowering, [e] → [n], opens the vowel, but keeps it fronted. (3) Depalatalization, [e] → [a], backs the vowel, but retains its height. The last two applied cumulatively sonorize the vowel completely to [a]. Finally, there is a process which splits the vowel into two halves, permitting each to polarize one of the two properties of the original single segment: (4) Diphthongization, [e] → [eq] or [æ], which develop ultimately to the optimal diphthongs [ai] or [æ] by the application of Raising (to which glides are especially susceptible), and of Lowering and Depalatalization (to which lax syllables are especially susceptible). It might be noted that once the essential properties of the vowel are understood, the existence and nature of these processes seems to follow almost necessarily from these. (The formulations here are due to Patricia Miller, revisions of Miller 1972a, 1972b, and Stampe 1972.)

In contrast to these context-free processes, which respond to the inner complexities of segments, there are context-sensitive processes, which respond to complexities of sequences of segments, e.g. vowel nasalization, which assimilates a vowel to a feature of its context. Sequences of segments often require radical and precise movements of the articulatory organs which are quite difficult in speech of normal tempos. There are processes that respond to these difficulties by substituting segments which 'blend' better into the sequence. A diphthong like [ai] requires a maximally open tongue aperture followed by a maximally narrow one. These extremes are mitigated by two processes affecting the syllabic on behalf of the nonsyllabic: (1) Palatal Assimilation, [ai] → [aï], and (2) Height Assimilation, [ai] → [aï]. Applied cumulatively they result in [æï]. In addition there is a process affecting the nonsyllabic on behalf of the syllabic: (3) Height Assimilation II, [ai] → [æg]. If all these processes apply cumulatively, the result is [æg]. This is subject to a final process, (4) Monophthongization, which eliminates the sequence [æg] altogether in favor of a single segment [æ].

The contrary teleologies of contrary processes are particularly clear in the contrary conditions under which they tend to apply. Context-free processes apply most generally in formal, legato speech. For example, diphthongization of vowels like [a] in Appalachian dialects is more frequent and extreme in slow or hyperarticulated speech. Context-sensitive processes, on the other hand, apply most generally in informal, legato speech. For example, in the
same Appalachian dialects the monophthongization of diphthongs like \([ai]\) to \([a] \) is more frequent in rapid or hypoarticulated speech.

These contrary teleologies also manifest themselves in the typical ordering of contrary processes in mature phonological systems. I noted earlier that in English the context-free process of vowel denasalization is ordered prior to the context-sensitive process of vowel nasalization. This order maximizes the paradigmatic distinctness of the underlying vowels of English (eliminating nasality in individual vowels), but minimizes their syntagmatic difficulty in connected speech (assimilating vowels to adjacent nasals). Processes which govern underlying phonological representations are predominantly (though by no means exclusively) context-free, while those which merely govern surface representations are predominantly context-sensitive.

Indeed, in casual speech, context-sensitive processes which normally govern only underlying representations in a language are often permitted to apply to surface representations. There is a process devoicing stops after tautosyllabic \([s]\), which is responsible for the absence of forms like \(/sg\in/\) in English, for our apprehension of the lax stop in \([sk\in]\) as a voiceless \(/k/\), for our pronunciation of foreign words like Italian \([sg\mathit{r}t\sim i]\) \(S\mathit{gr}t\sim i\) as \([sk\mathit{r}t\sim i]\). In informal speech this process applies to surface representations: if \(le\)’s go is reduced to \([sgo\mathit{y}]\), it normally undergoes devoicing to \([sko\mathit{y}]\). The study of casual speech turns up dozens of context-sensitive processes which extend their effects on underlying representation to superficial representation, but I have been unable to find a single context-free process which extends its effects in this way. For example, although vowel denasalization is a fully productive constraint on underlying representation in English (as can be seen in its application to nonderivable vowel nasality in foreign words) there is no casual speech style to which the process applies: \([k\mathit{\tilde{a}}t]\) \(c\mathit{an}\)’t does not become \([k\mathit{st}]\) no matter how rapidly or relaxedly one speaks.

The characteristic ordering of paradigmatically motivated processes prior to contrary syntagmatically motivated ones is responsible for an important constraint on underlying representation, that it may not be less underlying than the level of representation traditionally called phonemic. I have discussed the evidence for this constraint in some detail in Stampe 1968. One sort has been mentioned already: that English speakers identify all their vowels as nonnasal, even in forms which invariably are pronounced with nasalized vowels, like \([k\mathit{\tilde{a}}n]\) \(c\mathit{om}\) and \([m\mathit{\tilde{a}}t]\) \(m\mathit{int}\). Additional evidence comes from blends like \([m\mathit{\tilde{a}}\mathit{\tilde{u}}m\mathit{\tilde{n}}]\) or \([m\mathit{\tilde{a}}\mathit{\tilde{i}}\mathit{\tilde{t}}\mathit{\tilde{n}}]\) for \([m\mathit{\tilde{o}}\mathit{\tilde{u}}\mathit{\tilde{i}}\mathit{\tilde{t}}]\) mostly and \([m\mathit{\tilde{a}}\mathit{\tilde{u}}\mathit{\tilde{i}}\mathit{\tilde{n}}]\) mainly (also observed, but without phonetic detail, by Fromkin 1971:40), slips like \([f\mathit{\tilde{a}}\mathit{u}\mathit{t} \eta \mathit{e}\mathit{n}\mathit{d}]\), \([f\mathit{\tilde{a}}\mathit{u}\mathit{t} \eta \mathit{f\mathit{\tilde{a}}\mathit{\tilde{u}}\mathit{\tilde{a}}\mathit{\tilde{n}}\mathit{d}]\) \(l\mathit{o}\mathit{s}\) and \(\mathit{f}\mathit{o}\mathit{u}\mathit{n}\) (Fromkin 33), examples with nasalization blocked by resyllabication, as in singing \([d\mathit{z}\mathit{e}\mathit{\tilde{e}}\mathit{n}\mathit{\tilde{u}}.\mathit{f\mathit{\tilde{a}}}]\) for \([d\mathit{z}\mathit{e}\mathit{\tilde{e}}\mathit{n}\mathit{\tilde{u}}.\mathit{f\mathit{\tilde{a}}}]\) \(J\mathit{e}\mathit{n}\mathit{\tilde{e}}\mathit{f}e\) or \(P\mathit{ig} \mathit{L}\mathit{a}\mathit{t}i\) \([\mathit{g}\mathit{u}.\mathit{\mathit{\tilde{n}}\mathit{e}\mathit{m}]\) for \([\mathit{f}\mathit{\tilde{a}}\mathit{\tilde{u}}\mathit{\tilde{m}}\mathit{m}]\) \(\mathit{c\mathit{om}}\), etc. In all these examples an underlying nonnasal vowel surfaces when the process producing surface nasality is blocked. Note that we
cannot attribute this absence of nasality to a process denasalizing nasal vowels before nonnasals; clearly no such process exists in English, since the nasality of [kæt] can’t is never lost even though the vowel precedes a nonnasal in surface representation.

The explanation for the English speaker's analysis of invariably nasal vowels, as in [ɘʊm], as underlyingly nonnasal, /əʊm/, comes from corollaries of our theory. The innate denasalization process need not be overcome in language acquisition unless it confronts counterinstances. The nasalized vowels of the above words need not be counterinstances, since they can be derived from nonnasal vowels by innate nasalization processes which themselves confront no counterinstances in English. The denasalization process can be saved simply by ordering it prior to the nasalization processes, and taking all the superficially nasal vowels to be underlyingly nonnasal. (Of course, this would be impossible in a language—Hindi seems to be an example—in which all superficially nasal vowels cannot be derived through natural nasalization processes from nonnasal vowels; in such a language the denasalization process would have to be suppressed, and thus nasal vowels would be phonemic in the language.) The ordering of denasalization prior to nasalization by a child can be seen in Joan Velten’s developing speech (page 21 above). This ordering is encouraged by another corollary of our theory, that paradigmatically motivated processes are ordered prior to syntagmatically motivated processes like nasalization.12

This relationship between contrary processes gives a systematic account of the notion 'allophone' in traditional phonemics. An allophone is a sound which does not occur in underlying (phonemic) representation, but only in superficial (phonetic) representation, due to a context-sensitive 'allophonic' process. An allophonic process is any process, like vowel nasalization in English, which creates sounds which do not occur in underlying representation in the language (and which therefore does not neutralize any underlying distinction). In natural phonology, the nonoccurrence of certain sounds (nasal vowels) in underlying representation in a language is attributed to a process (vowel denasalization) in the phonological system of the language. Thus the notion 'allophonic' process translates as any process which gives rise to sounds eliminated by a prior, more general process in the system. Vowel nasalization in English is allophonic because it gives rise to sounds which the prior context-free process of vowel denasalization eliminates—nasal vowels. And nasal vowels are therefore 'allophonic': they are 'allophones' of nonnasal vowels, in English.13

To recapitulate, the nature of 'allophones' is systematically defined by relationships of processes, and their nonoccurrence in underlying representations is explained by a basic assumption of natural phonology: that processes are expressions of the language-innocent speech capacity, and they are not overcome by the language learner unless they confront counterinstances in the language he is learning. Vowel denasalization is not overcome in a language lacking nasal vowels altogether. Similarly, in a language in which all nasal vowels can be derived through nasalization processes, it
suffices to order denasalization prior to nasalization. The inability to produce nasal vowels as such need not be overcome provided such vowels are underlyingly represented as nonnasal. Thus the conservation of energy in language acquisition requires that if all nonnasal vowels can be derived from nonnasal vowels, they must be so derived. Consequently, we find no nasal vowels in underlying representations in English.

It would be possible to multiply examples of derivable features which do not occur in underlying representations in various languages, but I think it will suffice here to mention the opposite case: derivable features which do occur in underlying representations. A case which is fairly well known from recent discussion occurs in German, in which final obstruents are devoiced: [vok] ‘road’ beside [veg] ‘roads’. In Lithuanian Yiddish (Sapir 1915) the devoicing process ceased to apply, so that we have [veg] ‘road’ instead of [vok]. (On the reason for this process ceasing to apply, see Stampe 1969:453.) But in forms which did not have inflected alternants with voiced obstruents, e.g. [avek] ‘away’, the final voiceless segment remains. Why wasn’t [avok] analyzed as /avok/? It is not sufficient to point out that [avok] invariably had [k]; in English, [kɔn] can invariably has [ɔ] but it is analyzed as /ɔn/. The difference lies in the fact that in English there is a process (denasalization) which generally bars nasal vowels like [m] from underlying representation; whereas in German there was (and is) no process that generally bars voiceless obstruents like [k] from underlying representation. (In fact there surely exists no process which would have this effect, since voiceless obstruents are universal in phonological systems, even those of the youngest children.) English [ɑ] has to be analyzed as /ɑ/, but final [k] in German can be taken at face value as /k/.

The devoicing process in German is not allophonic, according either to the traditional definition or to our own. Accordingly, it is *morphophonemes*, to use the traditional term: it neutralizes an underlying distinction. For example, [bunta] ‘colorful ones’ and [bunda] ‘associations’ are neutralized in their inflected forms as [bunt].

The surfacing of final voiced obstruents after Yiddish suppressed the devoicing of final obstruents, e.g. in [veg] ‘road’ for previous [vok], plural [vegə], indicates that underlying representations of these words contained voiced obstruents, e.g. /veg/. Otherwise it would have been impossible for speakers to pronounce plurals like [vegə]; the voiced obstruent cannot be derived from a voiceless one, with contextually conditioned voicing, because there are words like [zeka] ‘bags’ which require any process which might be used to account for the voiced element to be suppressed.

Traditionally, notion ‘contrast’ would have been introduced to explain the /k/ versus /g/ representations. In natural phonology this results entirely, as in this paragraph, from the acquisition of the pronunciation of the words in question.

The view of underlying representation that emerges is one somewhat resembling that of Sapir or Bloomfield: underlying representations are basically phonemic, with ‘morphophonemes’ like
the German final voiced obstruents included in individual morphemes whose alternants require them. Assuming that the normal situation is for the underlying representation of a form to be identical with its surface representation, there are two constraints that drive representations deeper than the phonetic level. One is the general constraint affecting all forms, that owing to the coexistence of contrary processes in a system, 'allophonic' properties are barred from underlying representation. This has the general result of making all underlying representation at least as deep as the traditional 'phonemic' level. The other constraint is specific to particular forms: if a form has alternants whose pronunciation cannot be derived from the phonemic representation, its representation must be 'deepened' accordingly.

This view of underlying representation differs rather strikingly from that generally assumed in recent phonological theory, which has taken the basic level of representation to be 'systematic phonemic', as defined, for example, by Chomsky (1964). This level is defined as one in which 'redundant' features are unspecified. (This similarity to the structuralist view of the nature of the phoneme is not accidental, but probably inherent in the basically descriptive and language-particular emphasis of both structuralist and recent linguistic theory.) Thus instead of /k/ in [avek] away we would have a velar stop unspecified for voice (its voicing being predicted by the obstruent devoicing process), an 'archisegment' embracing /k, ɡ/. However, as is indicated by the invariably voiceless reflexes of such sounds when Yiddish suppressed the devoicing process, these neutralized segments do not derive from abstractions lacking any voice specification. Rather, their underlying representation is identical to their phonemic representation, e.g. /avek/.

An apparent impediment to the view I am proposing has been pointed out by Halle (1958): Russian has a voicing assimilation that substitutes [j] for [z] and [d] for [t] in certain voiced contexts; this is the sole source of [j], while [d] and [t] contrast in other positions. Halle argued that the distinction between these allophonic and morphophonemic substitutions could be captured only if the integrity of the voice assimilation process were denied. This would have been true in traditional phonemics, where the phonemic status of substitutions was taken to be a primitive. In natural phonology, however, it is a corollary of the interrelations of processes. We can assume that there is indeed just one voice assimilation process in Russian, but that one of its substitutes, [j], is barred from underlying representations by an earlier, more general process (I don't know which of several possible processes this happens to be in Russian, but its identity is not at issue), whereas another of the substitutes due to voice assimilation, [d], is not barred by any earlier process (since, whatever such a process might be, the acquisition of [d] in nonderivable contexts would have required speakers to suppress it).
Parallel situations occur in English. For example, we have a process that assimilates [n] to the point of articulation of a following stop. (It is optional across syllable boundaries: \([n.\, bed]\) or \([n.\, bed]\) \textit{in bed}, \([mn.\, ka\, s]\) or \([\text{w}\, n.\, ka\, s]\) \textit{one case}.)

The substitution of [m] for [n] is a morphophonemic one, there being no general process ruling out /m/ in English. On the other hand, the substitution of [\(\eta\)] for [n] is allophonic, since [\(\eta\)] occurs only where it arises by assimilation. The general process barring [\(\eta\)] from underlying representation is \([\eta] \rightarrow [n]\). This is evidenced by this substitution in foreign words with [\(\eta\)] which can't be derived from /n/, e.g. [\text{nuyen ftngen} by children's spellings (Read 1971), and so forth. In forms where no alternation occurs, e.g. [\text{lump}], [\text{lunt}], [\text{lunk}], we would therefore expect the underlying representations /\text{lump}/, /\text{lunt}/, /\text{lunk}/, since [\(\eta\)] is barred from underlying representation by the process mentioned above. Notice that these representations correspond precisely to English spelling. ('Systematic phonemic' theory treats these nasals as underlingly unspecified for point-of-articulation, markedness theory (Chomsky & Halle 1968; Chapter 9) treats them as /n/. Neither is capable of accounting for the spelling.) 16

Previous phonological theories have assumed that there is a general distinction between processes which govern underlying representation (these have variously been called phonotactic, morpheme structure, and redundancy rules) and those which govern derived representation (morphophonemic and allophonic rules, 'P-rules'), and that this distinction is arbitrarily determined in each language. But there are processes which in certain languages have both roles, like [n]-assimilation in English. There is no reason to believe that English has two such processes, one for each role, any more than there is reason to believe that Russian has two voice-assimilation processes. The distinction between the roles of processes is not an extrinsically imposed one. Rather it is determined by natural interrelations of processes in a given phonological system: any process (or, in the case of processes with dual roles, any subprocess) may govern underlying representation if its output is not barred from the lexicon by an earlier process which eliminates that output. Beyond this, the distinction is not general but morpheme-specific: a specific morpheme is represented phonemically unless its alternants are such that they can be derived in the system only from a deeper representation.

In support of this conclusion consider further the process that deletes [h] before nonsyllabic segments. This process has had a parallel progress in all the Germanic dialects it has appeared in, applying at first before less sonorant segments (e.g. older English [\text{nûtu]} nut, [\text{ðiaxn}] laugh, [\text{hring}] ring) and then before more sonorant ones (e.g. [\text{hwål}] why, and eventually [\text{hju\, \text{h}}] hew). Atomic words are more susceptible than tonic ones: there are dialects in which [\text{hwål}] whale retains [h] but atomic [\text{hwål}] why has lost it. This is sometimes lexical, so that even restressed why lacks [h].

These hierarchic conditions are reflected in modern English dialects. In Fig. 1 are represented four dialects I have observed
personally; the application of [h]-deletion is indicated as obligatory (+), optional (±), or inapplicable (−).

<table>
<thead>
<tr>
<th></th>
<th>atomic</th>
<th>tonic</th>
</tr>
</thead>
<tbody>
<tr>
<td>/n, l, r</td>
<td>/w</td>
<td>/w</td>
</tr>
<tr>
<td>A</td>
<td>+</td>
<td>±</td>
</tr>
<tr>
<td>B</td>
<td>+</td>
<td>±</td>
</tr>
<tr>
<td>C</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>D</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Fig. 1.—[h]-deletion in four English dialects (A-D)

The pattern of generalization of [h]-deletion from the most susceptible to the least susceptible contexts in these dialects leaves no doubt that we are dealing with a single process. All dialects have of course merged [hn, hl, hr] with [n, l, r], leaving no trace of the original [h] in alternations; therefore this much of the process governs lexical representation in all dialects.

In dialect A deletion is optional before [w] in atomic words; alternants like [hwai/wei] why require /hw/ in their underlying representation. In dialect B deletion is obligatory in these words, but only optional in tonic words; therefore it governs lexical representation in the case of atomic words, since they exhibit no alternants with [h]. In dialect C this lexical role is extended to tonic words, so that whales and whale, for example, are identically represented with /w/. In dialect D the process reaches its most general form, merging the lexical representations of hus and you with /j/.

In a theory which separates processes governing underlying representation from those governing derived representation, the connection between the nonoccurrence of /h/ before nonsyllabics of lesser sonority and its optional deletion before nonsyllabics of greater sonority would be missed altogether. Such a theory would be incapable of explaining why there are no English dialects which lack /h/ before /j/, for example, but only optionally delete /h/ before /w/. Furthermore, it would present the diachronic development from A to C as coincidental simultaneous generalizations of each of the two processes, and the development from C to D as a generalization of the process governing underlying representation and a coincidental simultaneous loss of the process governing superficial representation (since, if /h/ does not occur before nonsyllabics, there would be no reason to posit an [h]-deletion process governing superficial representation). But these are not separate coincidental developments. It is a single process that is generalized here. Its distinct roles in lexical and derived representation result simply from the fact that no phonological traces of /h/ remain when it is obligatorily deleted. There is no separation of processes with dual roles.

A corollary of the theory of underlying representation presented here is that underlying segments are ontologically of the same status as any segment in surface representation; they are mental representations of sounds which are, at least in principle, pronounceable. They are not, in particular, semi-abstractions like the 'archisegments' of structural and generative phonology. In
those theories, the \( [p] \) of spin is not identified with the voiceless \(/p/\) of pin or the voiced \(/b/\) of bin, because its voicelessness is not distinctive (there being no \( s\)bin) but predictable: obstruents are voiceless after \([s]\). Rather the \( [p] \) of spin is considered an 'archisegment', a labial stop \(/p/, b/\) unspecified as to voice.

Archisegments came into phonological theory after the mentalist view of underlying segments (phonèmes) of Baudouin de Courtenay (1895), Sapir (1933), and others was supplanted by the antimentalist view of Saussure (1959), the Prague school, Bloomfield (1933), and others, that phonemes are 'functional' entities characterized solely by those properties which distinguish them from other phonemes in a language, their 'distinctive' features. As Twaddell (1934) first recognized, this view of the phoneme as a set of 'distinctive features' leads to the conclusion that besides the distinctively voiceless \(/p/\) of pin and the distinctively voiceless \(/b/\) of bin, there is also the redundantly voiceless \(/p/, b/\) of spin.

But there is abundant evidence that the \( p'/s \) of spin and pin are phonologically identical, as opposed to the \( b \) of bin, and that in general stops after \(/s/\) are phonologically voiceless. There is the persistent orthographic tradition in English (and comparable languages) of writing \( p t k \) rather than \( b d g \) after \( s \). There are the spontaneous spellings of children with little knowledge of English orthography—WISPR whisperd, SCICHAT Scotch tape, MISTR monster, SKEEIG skiing, STARTID started (Read 1971). There are the voiceless stops which appear when \(/s/\) is transposed in slips of the tongue, e.g. \([h\text{wipsr}]\) (not \([h\text{wibsr}]\)) for whisper (Fromkin 1971). There is the fact that intensive \(/s/\) in crunch/srunch, trample/strample, mash/amash, etc. (a long list of English doublets occurs in Wright 1905:242-44) is never added to words with voiced stops: bash but not spash, grouch but not scroug, etc. 18

In the face of such evidence, it is remarkable that Twaddell's observation that the archisegmental view of the phoneme treats these phonologically identical segments as nonidentical did not lead to its abandonment. But the psychological facts were never sought out: neither Twaddell nor the phonologists who followed ever attempted to find empirical evidence for or against archisegmental representation. In one version or another, it was explicitly accepted by many phonologists—Firth (1935), Trubetzkoy (1936), Harris (1951), Jakobson et al. (1951), Halle (1954)—and ultimately it was incorporated as a basic characteristic of the 'systematic phonemic' theory of generative phonology (Halle 1962, Chomsky 1964). The revision of this theory to eliminate certain formal problems associated with unspecified features (Stanley 1967, Chomsky & Halle 1968: Chapter 8) does not eliminate the critical property of archisegments, since it leaves the identity of the segments indeterminate. As the evidence cited in the previous paragraph indicates, it is not indeterminate: archisegmental theories of representation are simply not correct.

In our theory the identification of the stop spin with that of pin rather than that of bin follows from the fact that there is
no process that eliminates voiceless stops affecting this post-
sibilant context, whereas there is a process eliminating post-
sibilant voiced stops in favor of voiceless ones. This was cited
earlier in the example [skoʊ] let’s go. This identification
depends, in other words, on the interrelation of processes in the
natural phonological system of English. This is undoubtedly the
correct principle of identification, since it is obvious from
the subtlety of the evidence cited in the last paragraph but one,
and from the generality of the identification (it is true of all
nonsibilant stops) that it is not learned. To summarize, segments
are taken at face value unless this is prevented by their alter-
nants or by a general process that bars them from underlying
representation. This means that all underlying segments are
fully specified as to their underlying phonetic features.

This conclusion rescues the traditional assumption that each
language has an inventory of phonemes which is at least indirectly
accessible to its speakers' consciousness (Sapir 1925). This
assumption is clearly critical to an understanding of how alphabets
are devised and used. As McCawley (1968:89) has pointed out,
archisegmental theories of underlying representation are unable
to identify these inventories: to /p/ of pin and /b/ of bin they
add the 'archisegment' /p, b/ of spin, to /m/ of limb and /n/ of lin-
seed they add the 'archisegment' /m, n/ of lint, and so forth.
But there are no alphabets which furnish graphic equivalents for
underlying archisegments rather than segments. It is necessary to
find some account of the segment inventories which underlie actual
archisegments in archisegmental phonology. Such a straightforward account:

the set of segments not ruled out by the context-free processes
of the language in question.

This restores empirical justification to the study of
phonological inventories inaugurated by Trubetzkoy ([1939]1968)
and Jakobson ([1941]1968), and carried further by Patricia Miller
in her study of vowel systems (1972a). As these studies indicate,
there are principles governing the inventories of phonemes in
languages which are independent of those governing their in-
vventories of surface sounds. Jakobson discovered what he called
universal laws of 'unilateral implication' in the stratified
structure of the phoneme inventory. According to Jakobson, the
occurrence of mid vowels implies the occurrence of corresponding
high vowels, the occurrence of front round vowels implies the
occurrence of front unround vowels, and so forth. These implica-
tional laws govern also the acquisition and change of phoneme
inventories; thus the acquisition of a mid vowel implies the
acquisition of the corresponding high vowel, and the loss of a
high vowel implies the loss of the corresponding mid vowel.

Miller's study assumes that phoneme systems are governed by
innate context-free phonological processes, and seeks to derive
Jakobson's implications laws from the nature and possible inter-
relations of these processes. For example, the law that front
round vowels presuppose front unround ones derives from a corre-
spanding process unrounding front vowels:

\[
\begin{array}{c}
V \\
\text{palatal} \\
\text{flower}
\end{array}
\]
(The process is exemplified in children (e.g. Leopold 1947) and in historical change, e.g. Middle English [fɪr] > [fɪr] fire, [dʒp] > [dʒp] deep. As is indicated by the notation [!lower], lower vowels are more susceptible to unrounding than higher ones, e.g. Old English [ɒks] oxen > [eks] while [fɪr] fire remained round. There is no context-free process rounding front vowels. Consequently, due to the unrounding process, the mastery of front round vowels implies the mastery of front unround vowels. The implicational law does not predict what will happen to front round vowels if they are not pronounceable by a child or adult. The process does, and since the implicational law can be derived from the process, it need not be considered a primitive of phonological theory.

Similar reasoning can be extended to other implicational laws. The law that mid vowels imply high vowels is true, as Miller shows, only for chromatic (palatal or round) vowels, and for such vowels it is also true that low vowels imply mid vowels. This corresponds to a process raising chromatic vowels one degree of height:

\[
\begin{align*}
\text{V} & \quad \left[ \begin{array}{c}
\text{palatal} \\
\text{round} \\
\text{lower}
\end{array} \right] \\
& \quad \rightarrow \text{higher}
\end{align*}
\]

(This process is also exemplified in children (Leopold 1947, Velten 1943) and in historical change, e.g. 15th century English [dʒp] > [dʒp] deep, [wʊk] > [wʊk] weak, [gʊs] > [gʊs] goose, [bʊn] > [bʊn] bone. The greater susceptibility of lower vowels to raising is illustrated by the optionality of 16th century English [wʊk] > [wʊk] weak while [wʊk] > [wʊk] wake obligatorily.) In this case the process and its hierarchy of applicability [!lower] are the basis for the corresponding implicational law.

It must be noted, however, that there are exceptions to this implicational law due to the interference of other processes. Diphthongization of high tense vowels, as in the English Vowel Shift, where [mʊs] > [mʊs] mouse, [mʊs] > [mʊs] mice, left a system with [ʊ, ʌ] but no [T, U]. (In Stampe 1972 I give evidence that this diphthongization preceded the raising of [ʊ, ʌ] by over a century.) In several American Indian languages, e.g. Yokuts and Southern Chehalis, and in Pashto and other languages long (presumably lax) high vowels are lowered to mid. In other words, the law that mid vowels imply high vowels is true just to the extent that the corresponding phonological process (in this case, raising) is not overridden by conflicting processes (diphthongization, lowering).

Ultimately, of course, the explanation of phoneme systems must go beyond the processes that underlie the system, to the phonetic causalities that underlie the processes. Jakobson attempted to make a direct leap from systems to phonetic causalities. The laws he proposed confront many exceptions. In Miller's work, phoneme systems are referred to systems of phonological processes, and these in turn to phonetic causalities. (I have sketched some examples of the causalities involved above on pages 22-24.) As we have seen, processes respond to phonetic difficulties in ways that are appropriate but which may conflict with other processes. The
inclusion of processes in the theory of phoneme systems enables Miller to account straightforwardly for many vowel systems which formerly seemed aberrant, and to discover many regularities of vowel systems which remain hidden from a static viewpoint such as Jakobson's, which does not allow for the natural competition among processes.20

This returns us to my remark that nature did not furnish a single, coherent system for dealing with the limitations of our speech capacity, but rather with a set of highly specific and often partially contradictory processes. Undoubtedly the cross-purposefulness of these processes is merely a reflex of conflicting characteristics of the capacity for speech itself, which, from a biological standpoint, is after all only secondarily for speech. The speaking animal is imperfectly adapted for speech. Furthermore, if we have Jakobsonian expectations, man is imperfectly adapted even for evolving sound-systems that minimize his imperfections. Recent attempts to discover, for systems of \( n \) vowels, the perceptually optimal systems (Lindblom 1972), can hardly cope with the existence of 3-vowel systems as diverse as \( i a u \) (Arabic, \( i a a \) (Abkhaz), and \( i e a \) (Hildegard Leopold). All of these are readily accommodated by Miller's process theory of vowel systems. Phonetic studies of perceptually and articulatorily optimal distinctions between sounds must be referred not to the phoneme systems themselves but rather to the processes that govern such systems.

CHAPTER II

THE ORGANIZATION OF PROCESSES

There has been a tendency in linguistics to view processes of grammar as descriptions of the language 'competence' of speakers, and not of the actual processes that occur in the production or perception ('performance') of speech. In certain aspects of linguistics this hedge is well-justified, but ultimately it cannot be satisfying. Explanations of linguistic phenomena based on 'as-if' descriptions remain 'as-if' explanations. Limitations of the competence/performance type are required by the complexity of language, which calls for a divide-and-conquer strategy. But such limitations presuppose that theories devised on the basis of the limited data will not be fundamentally inadequate when the data are expanded. This presupposition could hardly be less plausible than in the case of the competence/performance dichotomy, since it becomes increasingly clear that in general the conditions of the use of language (performance) are responsible for the nature of language. This is not by any means a peculiarity of phonology, but phonology presents a particularly striking case.

The theory I am presenting assumes that systems of phonological processes are real, that the underlying and superficial representations of utterances really exist, and that they are constrained and interrelated by the actual agency of these processes. It assumes that
when processes perform substitutions, these are actual substitutions occurring in the performance (mental as well as physical) of utterances. Without these assumptions, the theory would merely be a 'model', furnishing not literal but only analogical explanations.

The assumption of reality may appear doubtful given the evidence cited in Chapter I that phonological substitutions occur in the mind. But their reality is corroborated by a striking fact: that phonological processes apply to the result of 'slips of the tongue'. Thus in slips like [θət] skē[t] or [kθət] ste[t] for scotch tape [skæ[t] əθt], the aspiration of the stops and the palatality of the [k] are adjusted in accordance with their new contexts: we do not find slips pronounced [tæ[t] skæ[t]] or [kæ[t] əθt]. Similarly we have, for [spægɛɾɪ] spaghetti, [skæbɛɾɪ], [gæpɛɾɪ], etc., with aspiration, voicing, and palatality readjusted; [sɡæpɛɾɪ] and [ʃɡæpɛɾɪ] do not occur. Many further examples could be given here, but the observation is not a novel one; Fromkin 1971 cites various corroborative studies. Slips are scramblings of features, segments, or sequences which occur occasionally in the short-term memory storage of utterances in speech-production. Phonological processes apply after slips occur, and therefore must also apply in the processing of individual utterances.

How 'phonological constraints, when learned, become behavioral constraints which occur after [slips]' (Fromkin 41), is a puzzle only if we accept the assumption of most phonological theories that the constraints are learned rather than innate. The evidence of child language shows clearly, however, that phonological constraints are innate, and do, from the outset of speech, govern phonetic behavior. It would be surprising if these constraints did not persist in adult speech.

There are other phonological constraints and alternations, different in many ways from those involving processes, which demonstrably are learned. For example, the alternation of [k] with [s] and [g] with [dʒ] before reflexes of front vowels in words of Romance origin, e.g. electric [k] beside electricity [s], pedagogue [g] beside pedagogy [dʒ], is clearly learned. Children do not begin by pronouncing words like kitty and get with [s] and [dʒ], and then limit these substitutions just to Romance derivatives. Some words they merely learn with [s] and [dʒ] in the first place, but those which they create or learn from reading sometimes fail to show the alternation (one hears adults say pedago[g]y, fun[g]j). Now, the significant thing about learned alternations like this is that they are utterly insensitive to slips: the spoonerism of cynical guys is not [dʒɪŋkɪ] kæ[ʃ'], as we would expect if this learned rule applied after slips occur, but simply [ŋkɪŋ] sa[ʃ']. (Note that the palatality of [ŋ], which is governed by an innate process, is readjusted after the spoonerism occurs.) Thus phonological constraints which are learned do not govern our phonetic behavior.

The distinction between innate and acquired constraints is readily perceived by speakers. An innate constraint represents a palatable restriction on the speech output. It is quite difficult
for a monolingual English speaker to suspend the innate process governing aspiration of voiceless stops at the onset of stressed syllables; this is reflected in his treatment of the spoonerism [kʰætʃ step] for [skætʃ theip] scotch tape, in his aspiration of tonic stops in French words like [patʃ] petite, and in his failure to aspirate stops in words like Gurkha where the stress does not follow the stop. Even in the case of optional processes, there is perceptible difficulty. The process palatalizing apical obstruents before [j], as in [gæʃjo] beside [gæʃjo] got you, can certainly be suspended, but it requires some conscious effort to do so in actual discourse. On the other hand, to suspend the acquired obligatory substitution of /s/ for /k/ in electric/electricity presents no articulatory difficulty whatsoever, as is apparent in the facetious pronunciation electricity.

In the following pages I will discuss some characteristics of phonological processes and how they are organized into systems. In this discussion it is essential that the distinction between innate phonological processes and acquired phonological 'rules', like that governing the k/s alternation in English, be kept firmly in mind, for these two classes of constraints have quite different characteristics. Recent work in phonology has been marred by a failure to make the distinction between processes and rules, on the mistaken assumption that all constraints and substitutions are governed by acquired rules. As a result certain properties of processes have been attributed to rules and, conversely, certain properties of rules have been attributed to processes. For example, because many processes are optional, it has been supposed that rules also may be optional; in fact I do not know a single clear case of an optional acquired phonological rule. On the other hand, because certain rules make radical substitutions, like those in mouse/mice, wind/wound, it has sometimes been supposed that processes might likewise do so; in fact it appears that individual processes make minimal substitutions, and that seemingly contrary cases actually involve several processes applying in sequence. The distinction between processes and rules, as I understand it, is an absolute one, a distinction between constraints which the speaker brings to the language and constraints which the language brings to the speaker, whose distinct origins are reflected in their distinct roles in speech production.21

Another result of the failure to distinguish between processes and rules has been an uncritical willingness to attribute to processes cognitive characteristics which seem quite out of keeping with their infant origins. For example, Kiparsky (1968) notes that in Old English there were two contexts in which vowels were shortened: before three-consonant clusters (e.g. bræmbles 'brambles' from bræmbles), and before two-consonant clusters followed (roughly) by two unaccented syllables (e.g. blætsjan 'bless' from blætsjan). The latter is nowadays referred to as trisyllabic shortening (Chomsky and Halle 1968). On the assumption that Old English speakers had to learn these constraints, it is not unreasonable to assume further that since both involve shortening
of vowels, speakers would have generalized their common properties. Generative phonological theory furnishes a device for such combinations of rules:

\[
V \rightarrow \text{-long} / \_\_CC \left\{ \begin{array}{c}
C \\
\text{V(C)V}
\end{array} \right\}
\]

To show that this device represents a psychologically real way of combining the shortenings, Kiparsky points out that in early Middle English, both constraints simultaneously were tightened by one consonant:

\[
V \rightarrow \text{-long} / \_\_C \left\{ \begin{array}{c}
C \\
\text{V(C)V}
\end{array} \right\}
\]

E.g. *husband* > *husband*, *stirrups* > *stirrups*, *divinity* > *divinity*. By recognizing the reality of the bracket, Kiparsky argues, it is possible to explain why these otherwise unconnected constraints were changed simultaneously in the same way.

Granting the beauty of this argument, it nonetheless seems odd that phonetically motivated changes of this sort, changes which at least in their inception were imposed on the language by its speakers, and not vice versa, should be subject to the sort of cognitive analysis implied by the brackets. When this oddity is investigated more thoroughly, it turns out that in fact the changes have been incorrectly analyzed. The correct analysis makes shortening in closed syllables and in trisyllabic contexts a single process:

\[
V \rightarrow \text{-long} / \_\_CC.
\]

(. represents a syllable boundary.)

And for Middle English:

\[
V \rightarrow \text{-long} / \_C.
\]

This reanalysis is motivated by the failure of shortening to apply before Old English three-consonant clusters which were syllabicated C.CC, e.g. *hiehsta* 'highest', or before Middle English two-consonant clusters syllabicated .CC, e.g. *f\text{\textae}ponden*, *mitron*. As for the trisyllabic cases like *bleta\text{i}jan*, *divinity*, these typically involve a weakly accented medial vowel, as is shown by the tendency of this vowel to be syncopated both in Old and Middle English, e.g. *fad\text{\textae}r/fad\text{\textae}re\text{\text{"{e}}}s* 'father/father's' .

As in Modern English, a weak vowel is incapable of supporting a consonant in its syllable; compare Modern English [\text{\text{"{e}}}\text{\text{\textae}}\text{\text{"{e}}}\text{\text{"{e}}n}] zeroing (beside [\text{\text{"{e}}}\text{\text{\text{"{e}}}n}] in many dialects), [\text{\text{\text{"{e}}}\text{\text{\text{"{e}}}n}] \text{divinity} (with flapping of syllable-final [\text{\text{"{e}}}], contrast [\text{\text{\text{"{e}}}\text{\text{\text{"{e}}}n}] \text{\text{\text{"{e}}}n}] \text{\text{\text{"{e}}}\text{\text{\text{"{e}}}n}]). The weak medial vowel in the Old and Middle English examples likewise forced the preceding consonant into the syllable to its left: *bleta\text{i}jan*, *di\text{\text{"{e}}}\text{\text{\text{"{e}}}n}i\text{\text{"{e}}}ty*. Given this syllabication, trisyllabic words were shortened by the same closed-syllable condition responsible for the shortening of *br\text{\textae}mb\text{\text{"{e}}}s* and *hus\text{\text{"{e}}}band*. There is only the one process, responding to closed syllables; to explain the change from Old to Middle English no recourse to brackets is needed, because under the proposed analysis it is a simple change of the conditions for shortening from doubly- to singly-closed syllables. In eliminating the brackets in this case, we are eliminating the only example in which historical and psychological reality had been evidenced for brackets in a sound change.
Another cognitive device which might indeed play a role in the formulation of acquired phonological rules but which has been uncritically extended to processes is the alpha variable, which permits assimilation processes to be expressed as reciprocal changes (Chomsky and Halle 1968: Chapter 7). For example,

$$V \rightarrow \text{nasal} / \text{nonnasal}$$

would not only make vowels nasal before nasals but would also make them nonnasal before nonnasals. The cognitive elegance of this formulation notwithstanding, it seems to be an incorrect analysis of the principles governing nasalization in English and, apparently, in any language. In Chapter I, I argued that there is in fact no phonological process denasalizing vowels before nonnasals.

In languages where vowel nasality seems to be reciprocally determined by the nasality or nonnasality of a following segment (English is an apparent example), this appearance is due to the application of a process denasalizing all vowels followed by a context-sensitive process nasalizing vowels before nasals. The reciprocal analysis would make it impossible to explain why vowels with invariant superficial nasality (i.e. vowels invariably occurring before a tautosyllabic nasal) are phonologically perceived as nonnasal in English and similar languages, as the evidence cited in Chapter I clearly indicates. The analysis distinguishing between context-free denasalization and context-sensitive nasalization explains this phonological perception. Besides, there is unequivocal evidence against the reciprocal analysis: when segments are deleted in casual speech, vowels exposed thereby to nasalization are nasalized, e.g. \(\text{shutl they're} \rightarrow \text{shut} + \text{it's + r}'\), but when the same deletions appear nasal vowels to the hypothesized reciprocal denasalization, it does not occur, e.g. \([s\#\text{scot}] \text{scot} + [s\#\text{scot}] + \text{scot}\). It is impossible to discuss here the merits of the variable device in each case where it has been employed in recent phonological analysis. (The controversial variable analysis of the English Vowel Shift proposed by Chomsky and Halle (1968) I have attempted to dispose of in Stampe 1972.) What I wish to emphasize here is that however plausible the device may seem as a strategy in the mental formulation of acquired rules, e.g. the rule proposed by Chomsky and Halle (1968) to describe the formation of the imperfect in West Semitic languages, there is no reason to expect logical devices of this sort to be imposed on constraints reflecting innate limitations of the speech capacity.

My third example of the use of cognitive devices appropriate to rules in the analysis of processes concerns a recent proposal regarding the ordering of contradictory processes. One of the processes involved is the trisyllabic shortening process, which I have just argued didn't exist. S. Anderson (1969: 137-43) cites this process

$$V \rightarrow \text{long} / \text{CV(C)V}$$

and another process of Middle English which lengthened stressed vowels in open syllables, e.g. \(\text{baken} \text{OE bacaen}\), etc. The vowels so lengthened were lowered: \(\text{wik/s\#kes 'week/weeks', sun/s\#nes 'son/sons'}\). Anderson's formulation, simplified slightly for this discussion, is:

$$\left[ V \right]^{+\text{long}} / \text{-high} / \text{CV}$$
He notes that if both these processes, as formulated, apply to a single form, they give incorrect results regardless of the order in which they apply. For Old English ival, Middle English had ival/iveles 'evil/of evil', and for somor, it had somor/sumeres. These forms do not result from either order of application:

- Base form: ival-es, sumer-es
- Shortening: x
- Lengthening: *iveles, *sumeres

Anderson notes that as formulated the context of the lengthening process CV properly includes the context of the shortening process CV CV. Citing a parallel principle of Pāṇini's grammar, Anderson proposes that when two processes contradict each other, the specific process applies and the general one does not. Thus in the three-syllable examples cited the lengthening process would not apply and the correct forms ival-es and sumer-es would result.

This makes sense as a hypothesis about how a language learner, or a grammarian like Pāṇini would formulate rules; obviously there would be no point in positing a specific rule if its effects were totally eradicated by a more general rule. But we are not speaking here of rules which are imposed on learners by the nature of the language. These are phonological changes, processes imposed on the language by its learners. And it is not at all clear why the proposed disjunctive principle should be applicable to processes. In any event the principle seems to be falsified, at least for processes, by the way that allophonic properties of sounds are ignored in underlying representation. As was shown in Chapter I, all English vowels are underlyingly nonnasal. This is fully explainable if we assume that the general process denasalizing vowels regardless of context governs representations up to the point that the specific process nasalizing vowels before nasals applies. But the disjunctive condition Anderson proposes would predict that denasalization would not apply to any vowel to which nasalization would later apply, and would make the fact that such vowels are phonologically perceived as nonnasal inexplicable.

But what about the two processes of Middle English? Again, the analysis seems to be faulty. I have already pointed out that the correct context for the shortening process was /_C_, and that it applied to trisyllabic forms because of their special syllabication, e.g. in.les, sum.eres, which was in turn due to the weakness of the medial syllable; this is corroborated by the fact that these forms have syncopated variants in Middle English, e.g. ives, sumres. The correct context for the lengthening rule, on the other hand, was /___, as in sun/sumer 'son/sons'. This is confirmed by words in which vowels were lengthened before two-consonant clusters, e.g. patron, when these clusters were syllabicated with the following syllable. It is also confirmed by the nonoccurrence of lengthening in words like citi, mogy, etc., where the short final [i] vowel apparently was too weak to carry the preceding consonant and forced it into the syllable to the left, thus closing the syllable so that lengthening did not apply. The same syllabizations of these words persist in Modern English.
[\text{pê}.\text{trn}] \text{note syllable-initial [t] instead of syllable-final [z]}, [\text{sir}.\text{i}] [\text{mê}.\text{i}] \text{note the obligatory flapping typical of syllable-final intervocalic [t] and [n], and the obligatory nasalization of a vowel preceding a tautosyllabic nasal}. \text{The contexts of shortening (/C/)} \text{and lengthening (/_,_)} \text{as corrected, are mutually exclusive. The problem of why lengthening did not apply to inless or sumeres is solved without appealing to the problematic principle of disjunctive application.}

I would like to turn to some characteristics of processes that do seem real, and to show that many of these can be understood as reflections of their phonetic functions. Consider, for example, the question of whether processes apply simultaneously or in sequence. It seems to be generally accepted that most if not all, of the processes in a system apply in sequence. The arguments that have been given for this are not as conclusive as they could be. The most conclusive evidence I am aware of involves the application of a process to the output of an optional process. Cases like this are fairly commonplace, and the dependence of the second process on the application or nonapplication of the first establishes their sequential operation beyond a doubt. Consider the process of English I have discussed which nasalizes vowels before nasal segments obligatorily, e.g. [\text{sēn}] \text{seen beside [sē] see}. \text{There is an optional process that deletes flaps, e.g. [\text{kēt}] kitty, optionally [\text{kēt}]} \text{If flap deletion applies before a nasalized segment, as in \text[hēt}] \text{hitting, the preceding vowel is nasalized: [\text{hēt}]}. \text{This vowel remains nonnasal if the flap is not deleted; [\text{hēt}] does not occur. The same is true of the output of an optional process deleting glottals: [\text{bēt}] button (not [\text{bēt}]), optionally [\text{bēt}] (not [\text{bēt}]). And of the output of an optional process nasalizing voiced consonants before nasals: [\text{wēt}] wooden (not [\text{wēt}]), optionally [\text{wēt}] (not [\text{wēt}]). In each of these cases, the application of the nasalization process is strictly dependent on the application or nonapplication of a separate and unrelated optional process, and the obvious explanation for this dependence is that nasalization applies to the output of the other process, in sequence.}

Why should processes apply to the output of other processes? It seems obvious that simultaneous application could be much more rapid. We might conjecture that the neurology of speech processing is linear, and that processes are strung out one after another like consecutive operations on an assembly line. In generative phonology (Chomsky and Halle 1968: Chapter 8) it has in fact been assumed that processes are linearly ordered. The assumption has been challenged, however, by S. Anderson (1969), who argues that in Icelandic the processes of (1) syncope and (2) labial umlaut applied in (1-2) sequence in some words, e.g. \text{hamn}-\text{um} 'to the hammers' \text{– hamsun – hömsun}, and in (2-1) sequence in others, e.g. \text{baggl}-\text{t} 'to the package' \text{– boggul – böggi}. This conclusion casts doubt on the possibility of explaining the sequential application of processes as due to linear neural structure. 22

Before I propose an alternative explanation, I would like to give an example, unfortunately rather complex, that throws somewhat
clearer light on the nature of the nonlinear application of processes than do the examples of Anderson and others. The example is chosen from American English so that it can be readily confirmed by the reader. It involves several processes, of which three require some preliminary discussion: syllabication, flapping, and flap-deletion.

Syllabication attaches a nonsyllabic to the syllable to its right; but if the syllable to its right is unstressed and the syllable to its left is stressed, the nonsyllabic is instead attached to the syllable to its left. (This is the process responsible for the syllabication discussed above in Old and Middle English.) The process is subject to various constraints, which are not relevant to the example at hand, involving morphological boundaries and constraints on permissible syllable structure.


Since syllabication has not received the attention it deserves from phonologists or phoneticians, I should perhaps digress briefly here to mention some of the phonetic and phonological evidence for the existence and location of syllable boundaries. Syllable boundaries are not difficult to 'hear', once one has become conscious of the relevant cues. The chief phonetic cues, which are verifiable in spectrography, are the modifications ('transitions') that vowel formants undergo in tautosyllabic contact with various nonsyllabics.

In the syllabication V.CV these modifications occur at the onset of the vowel to the right; in VC.V, at the offset of the vowel to the left. The phonological evidence comes from processes whose application is dependent on the placement of syllable boundaries. There are many examples in English, some of which will be alluded to in the example to which this discussion is leading. Others include palatalization of velars in tautosyllabic contact with a palatal vowel, e.g. [by.kh,ří] okay versus [by.kř] Okay, and [ří.gó] eagle versus [ří.góy] ego; aspiration, which applies to voiceless stops only if they are the first segment in a stressed syllable, e.g. [dí.s.třá] distaste versus [má.střá] mistake, parallel to [dí.s.třá] this table versus [má.střá] the stable.

Syllabication is a process in the full sense, as is shown by the fact that in addition to establishing basic syllable divisions it can optionally alter them. For example, in casual speech the morphologically constrained syllabifications of distaste and this table may be relaxed to dis.taste and thi.stable, with e accordingly remaining unaspirated. More strikingly, syllabication optionally applies to unstressed syllabics, attaching them (with simultaneous desyllabification) to an adjacent syllable exactly as it attaches nonsyllabics to an adjacent syllable. For example, [ří.ří] real, optionally [říří]; [snó.ří] snowy, optionally [snoří]; [ří.kří] electric, optionally [ří.kří]. The attachment and desyllabification must be one simultaneous change. Otherwise intermediate stages would arise with n-syllable words containing n+1 syllabics (e*říří) prior to desyllabification, or n-1 syllabics (e*říří)
prior to attachment. These intermediate representations are universally unpronounceable, and there is neither evidence that such representations ever occur in phonological derivations nor reason to believe, given that the function of processes is to eliminate phonetic difficulties, that they could occur. Therefore it seems clear that optional desyllabification is carried out by the same process that establishes basic syllabifications, since the attachments of desyllabified segments obey precisely the same principles as the attachments of underlyingly nonsyllabic segments.

The second process to be discussed is flapping, which changes released apical stops [t, d, n] to the corresponding flaps after vowels (including [r], nonapical [t], [q], etc.) in syllable-final position. Before a following voiced segment [r] is obligatorily voiced: [hóthóus], with h-deletion [hár.úys]; [rát], [rárt] raty. Examples: [bár.ót] batting, [sár.át] sadder, [oúr.ú] thinner, [hár.ú] hearty, [wetr.ót] welded, [métr.ú] mental. The process is obligatory in most American idioms. Some speakers have optional syllable-initial applications before syllabics, e.g. [sá.rákt.ú] the doctor, [ví.rát] veto.

The final process to be discussed is flap-deletion, which optionally elides flaps in syllable-final position. For example, [bár.ót] batting, optionally [sá.rót]; [sár.át] sadder, optionally [sá.át]; [oúr.ú] thinner, optionally [oú.ú]. Flap-deletion does not apply to flaps in syllable-initial position, e.g. those cited at the end of the previous paragraph.

Now we are ready for the example, the phrase divinity judge. In casual speech it can be reduced to something like [dá.vít.i.tidʒ], and its derivation involves the application and re-application of the three processes discussed above, interwoven with each other and with other processes. To abbreviate the discussion, the derivation is taken up at an intermediate stage, to which vowel reduction has already applied. The asterisks mark forms which are unpronounceable, because there are obligatory substitutions which have not applied. Unstarred forms are pronounceable options.

0. other processes *dá.vínt.i tidʒ
1. syllabication: *dá.vínt.i.ti tidʒ
2. flapping: *dá.vínt.i.ti tidʒ
3. vowel-nasalization: dá.vínt.i.ti tidʒ
4. flap-deletion: dá.vínt.i.ti tidʒ
5. syllabication: *dá.vínt.i.ti tidʒ
6. vowel-nasalization: dá.vínt.i.ti tidʒ
7. a-harmony: dá.vínt.i.ti tidʒ
8. shortening: dá.vínt.i.ti tidʒ
9. syllabication: *dá.vínt.i.ti tidʒ
10. flapping: dá.vínt.i.ti tidʒ
11. flap-nasalization: dá.vínt.i.ti tidʒ
12. flap-deletion: dá.vínt.i.ti tidʒ
13. syllabication: *dá.vínt.i.ti tidʒ
14. vowel-nasalization: dá.vínt.i.ti tidʒ

This derivation does not exhaust the possible pronunciations of the phrase, nor is it the most extreme reduction possible, but it
should suffice to illustrate the way the processes interact with each other. They do not apply in a linear order, but rather apply and re-apply whenever the configurations they would eliminate arise. At 1 syllabication puts [n] into a flapping context. Its flapping at 2 produces a representation to which flap-deletion can apply, at 4. Flap-deletion, by eliminating the relatively obstructant flap between the stressed vowel and the unstressed [o], produces a representation vulnerable to (re)syllabication at 5. The third application of the syllabication process, at 9, leads to another application of flapping, this time to [t], at 10. [t] could have flapped optionally at any time, but it is not deletable until it becomes syllable-final at 9 (pronunciations like [de.vi.z.a.i], with syllable-initial flap-deletion, do not occur), and it could not become syllable-final until the resyllabication of [a] at 5 put [t] in immediately post-tonic position. Flap-deletion at 12 again removes an obstacle to resyllabication, which occurs once more at 13.

This example, which is paralleled by many casual-speech derivations in English involving these and other processes, shows clearly the random nonlinear sequential way that processes apply. (I will return below to cases where an ordering is imposed on the sequence of application.) In the example there is only one thing that seems to indicate nonrandom application: if flap-deletion were to apply immediately to the output of flapping at 2, the result would seem to be the nonoccurring *[de.vi.o.ti], lacking nasality. However, this is only apparent. Flap-deletion does not eliminate nasality, but only the flap gesture performed by the tongue; this is true of flap-deletion and other processes deleting segments not only in English but also in other languages in which they occur. Thus if flap-deletion applied prior to vowel-nasalization, it would leave nasality behind which would be extended over the vowel by the subsequent application of vowel-nasalization. This completely random order of application is particularly characteristic of optional substitutions, as in casual or rapid speech.25

Furthermore, the example establishes that processes apply iteratively. In Anderson's Icelandic example two processes apply in one order in some forms and in the other in other forms, but neither process re-applies. The divinity fudge derivation shows that re-application is possible. There are no fewer than four distinct applications of syllabication and two each of flapping and flap-deletion. It is this freedom to re-apply that enables us to characterize the application of these processes as a random sequence.

To understand why processes should apply like this it is sufficient, I think, to understand their respective functions. Syllabication determines the phrasing of 'sonants' (to use the classical term) and 'con-sonants'. It depends in part on stress, and therefore it stands to reason that accent should be established before syllabication occurs. (There are some mutual dependencies between stress-assignment and syllabication too complex to explore here, but the principle of priority suggested here seems capable of resolving these.) On syllabication, in turn, a very large number of processes depend. In the above derivation, for example, every
Process is at least partially dependent on syllabication: flapping is obligatory syllable-finally and optional otherwise, and vowel-nasalization is obligatory within syllables and otherwise optional (e.g. [sₕ˔ழ] sigur, [sₕ˔ʝNI] sigweur, [sₕ˔n˔] or [sₕ˔n˔s] Sînzi). Flap-deletion occurs only syllable-finally, e-harmony is nearly obligatory within syllables, and shortening occurs only in interior syllables (thus it occurs at 8 but not after 14). It would be possible to show why, given the phonetic functions of each of these processes, they depend as they do on syllabication. However, this may be left for the reader to divert himself with. The point is that they do depend on syllabication, and this dependency seems sufficient to explain why syllabication applies before they do.

But the re-applications of syllabication depend in turn on the application of segmental processes like flap-deletion. This follows from the fact that syllabication is restricted by the condition that sonority must increase from the margins to the peak (syllabic) of a syllable. Words like [aʃrnz] from illustrate this condition: if any two nonsyllabics are transposed the syllable is impossible, and must be resyllabified to be pronounceable; for example, *[aʃrnz], *[aʃrnz], *[aʃrnz]. In [doɔr.ə.ti] desyllabication of [ə] is impossible because it is flanked by less sonorant segments and cannot be attached to either. But when flap-deletion eliminates [ɹ], [ə] can be resyllabified. Thus syllabication, on which flap-deletion depends in the first place, depends in turn on flap-deletion. The sequence of applications and re-applications is completely natural, given the functions of the processes.

We do not have to appeal to speculations about neural organization to explain why processes naturally apply in sequence. Given the specific function of individual processes, simultaneous application would result in some processes creating the very configurations other processes are eliminating. This is true of context-free processes as well as context-conditioned ones.

In Chapter I, we saw that Joan Velten changed [ʒ] to [z], [j] to [ʒ], and [l] to [ʃ]. If these processes had applied simultaneously, she would have ended up having to pronounce two-thirds of the sounds the processes are supposed to eliminate. Instead, they applied in the natural sequence:

- adult form: lam lamb jard yard zak Jacques (1)
- delateralization: jam - -
- spirantization: jäm jard -
- depalatalization: zam zard zak
- other processes: zab zad zak

The same logic enables us to understand why some processes apply to strings of segments. Nasalization, which nasalizes sonorants prior to nasal segments, obligatorily if they are tautosyllabic and optionally otherwise, is a typical example. In a word like borrowing, which has the alternative syllabifications [b̩r.ow.ing] or [b̩r.o.ing]—the latter due to resyllabication of the glide of [ow], derived from [ɔ]], the interplay of the conditions of the nasalization process with the alternative syllabifications results in a complex pattern of admissible and inadmissible pronunciations:
As in the divinity fudge derivation, each representation in which no obligatory applications remain unapplied is pronounceable; these are indicated by the lack of an asterisk. The correct forms result quite straightforwardly from the iteration of the nasalization process, i.e., from its unrestricted application to any representation containing a sonorant prior to a nasalized segment. And since this is the sequence of sounds the process exists to alter, it stands to reason that it will apply not only to that sequence in underlying representation, but also in derived representations—even when the sequence arises due to the application of the selfsame process.

To my knowledge this explanation of the sequential application of processes, and of their nonlinear, randomly-ordered, iterative application, has not previously been proposed. The explanation is so straightforward that one may wonder why this should be the case. I think the reason lies in the assumption that substitutions are governed by rules which the language learner—the child—formulates to account for the sound patterns of the language he is learning. Since it is in principle possible to formulate any sequence of substitutions as a simultaneously applied rule, the rule hypothesis is inherently incapable of explaining sequential application. To be sure, the simultaneously applicable rule may have to include the conditions of all other rules on which its application depends. But this observation, which is commonly cited to justify sequential application, merely describes the problem; it doesn't explain why the sound patterns of languages should be such that they are best described as resulting from sequential substitutions. Besides, it isn't always true that simultaneous formulations of processes are more complex. The sequence of substitutions cited above from Joan Velten's speech could be formulated as a single rule making voiced oral apical continuants [l, j, 3] simultaneously into [z]. But despite its simplicity, this simultaneous formulation is incorrect. The constituent substitutions are separate and distinct, as is shown by their separate disappearance from a child's speech as it matures, and they occur in sequence, as is shown by the fact that the change of [l] to [z] entails a change of [j] and [3] to [z]. Indeed, the fact that the separate sequenced processes can be explained, along the lines proposed in Chapter I, whereas the generalized simultaneous rule cannot, is itself evidence that the simultaneous formulation is wrong. And given the distinctness of these processes, with distinct functions predetermined by distinct limitations of the innate speech capacity, the reason they should naturally apply in a random iterative sequence is almost self-evident.

In sequential applications such that the output of one process A creates input for another process B, A is said to 'feed' B, and AB is called a feeding order (Kiparsky 1968). The opposite order, BA
however, is the only case in which the term 'order' is justified, since, as the derivation of divinity fudge shows, so-called 'feeding order' is an iterative sequential application of processes in random order, no order at all. The derivation terminates when all applicable processes (plus or minus optional ones) have applied. It does not matter which of such a pair of processes as A and B applies first: AB is equivalent to BAB. It only matters, given B's function of eliminating representations of the sort A introduces, that B is allowed to apply to A's output. Therefore the ordering constraint or, better, an anti-sequential constraint which prevents B from applying to A's output can be viewed as unnatural.

However, such a constraint is occasionally met with in language. For example, in the dialect common to many speakers of the southern midwest, the vowel [e] is raised to [i] before nasal segments, e.g. [hin] hem, him, [tin] ten, tin. This process applies to certain sequences of [e] plus nasal which arise by rule, e.g. [mënt] for [mënt] meant, shortened from [mënt] mean, or [spëriti] for [spëriti], shortened from [spëriti] serene. (This is in accordance with the fact that processes generally apply after rules, as evidenced by slips of the tongue.) But the process does not apply to the same sequences when they arise from various other processes. Thus when words like [bëtëm] betting, [bëtëp] threaten, [rëdp] reddem become [bëtëm], [bëtëp], [rëdp] through the application of various consonant-deletion processes, they do not undergo a further change to [bëtëm], [bëtëp], or [rëdp]. In other words, the raising process is constrained not to apply to the output of these other processes. It is not difficult to establish, however, that the raising is learned in children who speak the dialect in question.
gested that such changes originate in the failure of children to impose constraints found in the language they are learning (Stampe 1969). It should be pointed out, however, that in the case of context-conditioned processes such changes are apt to be manifest only in the speech of older children and adults, for they are often heard only in fluent casual speech. Individual variation in the degree of reduction of various words and phrases is quite marked even in speakers of similar dialects, as becomes apparent in any detailed comparison of casual pronunciations. All midwestern speakers flap [n] (e.g. [fæn] funny), degeminate [nn] in Italian or Finnish words, and nasalize [d] after [n] in casual speech (e.g. [hänn] hand). Whether they apply these in unconstrained sequence, so that [kændi] candy + [känni] + [känl] + [känl], is a matter largely of individual style.

There are of course many other potential relationships between processes than the sort I have discussed. One of these involves processes which apply to partially identical configurations, so that whichever applies first eliminates some or all of the configurations to which the other might apply. Where process A eliminates potential input to process B, A is said to 'bleed' B. Bleeding order is generally characterized by Kiparsky (1968) as unnatural, but Kenstowicz and Kisseberth (1970) argue that bleeding of the context of another process may be natural. From a strictly phonological point of view I am doubtful whether there is any preferred ordering. I have found no consistent tendency in children's substitutions regarding processes in these relationships, and it is difficult to see why, if two processes both eliminate a certain phonetic configuration, it should matter for phonetic purposes which process accomplishes this, as long as one of them does. If there is indeed a preferred or natural order in such relationships, I suspect it may be motivated by other than phonetic considerations. In fact Kiparsky (1968, 1971) has emphasized the fact that these relationships may introduce irregularities in morphological paradigms.

The insight into the natural organization of processes which is provided by the study of their various teleologies is particularly striking in the case of processes which introduce opposite changes, such as the example of vowel denasalization and vowel nasalization discussed in Chapter I. The paradigmatic functions of context-free processes against the syntagmatic functions of context-conditioned processes require that they be ordered as given. Otherwise, their respective functions—maximizing distinctiveness and minimizing sequential difficulties—would not be realized. Thus it is that in Fox and many other languages all obstruents are voiceless (and thus maximally distinct from sonorants, which are voiced), but between vowels some obstruents are superficially voiced (and thus make the transition between vowels, which are voiced, less difficult). The opposite ordering would be pointless. Similarly, in early Germanic all round vowels were nonpalatal, but before high front vowels round vowels were palatalized (umlauted); in many languages all nasals are anterior, but before velars they become velar; in Sanskrit nonhigh vowels were nonlabial and nonpalatal, but in combination with labial and palatal glides within the syllable they became [eː] and [oː]; and so forth. It is this ordering principle that is
responsible for the distinction that exists between 'phonemes' and 'allophones', as discussed in Chapter I.

There is no reason to believe that we know all, or even most, of the principles whereby phonological processes are organized into coherent systems. What I have tried to show here is that some of the better-evidenced principles can be explained in terms of the natural functions of those processes. In fact I doubt whether we can understand the nature and organization of phonological systems without asking why such systems exist in the first place. Our earliest answers to such a basic question will surely be confused and riddled with error. But if they lead to better answers, or even to better-formulated questions, they cannot be dismissed as idle speculation.

REFERENCES


Stampe, David. 1968. Yes, Virginia... Unpublished paper read to the Fourth Regional Meeting of the Chicago Linguistic Society.


AFTERTHOUGHTS

To CHAPTER I:

1 (page 3) Perhaps Hildegard's "whispered" pronunciation of pretty was not exceptional after all. It may simply have been the regular reflex of an early process devoting high vowels in syllables with voiceless consonant, rather as in Portuguese, Japanese, etc. When she later learned to voice the vowels, the consonants would have assimilated to them in voicing, regularly, and if the R Leopold transcribed in the whispered pronunciation was, as I suspect, a natural misperception on his part of the stop release, it would have seemed to disappear in the voiced pronunciation. This example has received much attention (perhaps undeserved) in recent literature.

2 (page 16) This is incomplete. Since [t] might become [b] via either [d] or [b], the entailment is rather that either [d] or [b] (or both) would also become [b]. Donegan (1978) has discussed this categorial property of phonological derivations in more detail, and with far richer illustrations.

3 (page 18) I am alluding to the popular formulation

\[ V + [\text{nasal}] /\_ [\text{nasal}] \]

Usually, opposite changes reflect distinct processes with distinct telologies. As such, the application of one in a language never entails the application of the other. There are languages with context-free denasalization but no contextual nasalization (Japanese, in some descriptions), and languages with contextual nasalization but no context-free denasalization (Hindi). Changes due to a single process, on the other hand, are normally in a relation of unilateral entailment. As noted on page 8, if [i] is denasalized to [e], (it occurs) will also be denasalized to [e] under like circumstances. (See Donegan 1978 for numerous illustrations.)

4 (page 19) Thus languages which apply vowel denasalization lack nasal vowel phonemes, and also nasal vowel allophones (as in Japanese, in some descriptions) unless (as in English) assimilative vowel nasalization also applies. On the other hand, languages which suppress vowel denasalization have nasal vowel phonemes; such languages also suppress assimilative vowel nasalization (as French does, compare [bon] bonne), or they may apply it (as Hindi does) and have some nonnasal as well as nasal vowel phonemes realized as nasal vowel allophones. For every pair of contrary processes of this sort, there are just four possible phonological typologies (assuming that if the context-free process applies it must apply before the context-sensitive process.)
features ever play a role in phonological processes. Since it is by now well established that nonadjacent features often condition changes in distinctive features, the way is open to interpret actions at a distance as involving auditorily unobtrusive overappings of harmonized vowel features on intervening consonants, and likewise of harmonized consonant features (e.g. retroflexion in Sanskrit) on intervening vowels. This would explain why actions at a distance entail the corresponding actions in adjacent segments. E.g. Old High German palatal umlaut as in manni > men(i) 'men' entails stain > stein 'stone'; we would posit an i-colored nn in manni. This is a recurrent hypothesis in historical phonology, and it has been given up too readily in the face of surprising examples. I see no reason why even consonant harmonies like [pibl] r.v. or [gogl] doggle in child speech could not be explained in this way; it is the timbre, not the stoppedness, of the consonants which would be posited on the intervening vowels, and I see no a priori reason why the adjustments of lip and tongue musculature accompanying obstructions might not co-occur, unobtrusively to the ear, with the rather different adjustments accompanying vowels.

A simpler hypothesis regarding [hua] home beside [ham] ham would be that Joan pronounced [hu] as [hu], as e.g. in Japanese, since [g] would not be nasalized and would therefore cause progressive denasalization. Velten's phonemic analysis does not give phonetic detail, and the notes on which it was based seem no longer to exist.

The application of context-free vowel nasalization as (6) after context-sensitive nasahzation (1) is a counterexample to the precedence constraint proposed between theem types of processes on page 23f. I suspect that Joan's actual pronunciations were as in (5) rather than as in (6); Velten would simply have ignored vowel nasalization where it was obviously contextually conditioned, as in (5). Joan did denasalize vowels, as in [na] for French non (specifically remarked) and [atsu] for [atsu] ants, but there is no reason this could not have been prior to vowel nasalization at (1). This eliminates the counterexample.

Left out of this account are processes of Palatalization (the opposite of Depalatalization), Tensing, and its opposite, Laxing. Diphthongization can be eliminated, being simply an application of one of the other processes to half of a vowel. The vowel research of Donegan (formerly Miller) is brought together in Donegan 1978. The assemental hypothesis of Donegan and Stampe 1979, and Donegan 1978, according to which processes have only prosodic domains (e.g. syllables, or syllable constituents; measures, or measure constituents, etc.), and never segmental ones, carries with it a natural segmentation hypothesis that adjacent phonetic segments are distinct only in case they differ phonetically (which is dubious in the case of [eg] or [ge]) or occur in distinct prosodic constituents (e.g. geminate consonants on either side of a syllable division).

Donegan 1978 shows clearly that monophthongization is not a separate process, but simply the end product of the mutual assimilation of the features of a diphthong. This is in accordance with the natural segmentation hypothesis sketched in note 7.

The applicational precedence of context-free processes before contrary context-sensitive processes seems to be a special case of a much more general constraint Donegan and I (1979) have presented as "Fortition first, lenition last".

Some further discussion and evidence has now been published in section 4 of Donegan and Stampe 1979.

The re-analysis of the place of denasalization in Joan's speech proposed above in note 7 is consistent with an absolute interpretation of this constraint. Donegan and I (1979: section 3) have argued for such an interpretation, extended from pairs of contrary processes to all paradigmatic and syntagmatic processes ("Fortitions first, lenitions last"). For example, a fortition like the stress vowel in [bared] bared rather than as in [beted] bated could not be derived (through these processes). This constraint is extremely restrictive, barring analyses of a wide variety, including those Kiparsky sought to eliminate in his 'How abstract is phonology?' paper of 1968, and explaining why alternations of the divine : divinity type (where historically, the syntagmatic shortening of the stressed vowel in triplet words preceded the paradigmatic vowel shift and diphthongization) cannot, despite Chomsky and Halle 1968, be part of the synchronic natural phonology of a language. For further discussion see the notes on 'bleeding order' under Chapter II.

The characterization of the difference between phonemes and allophones in structural phonology in terms of criteria like 'contrast', 'complementary distribution', 'phonetic similarity', and so forth has been effectively criticized by Chomsky 1964, Postal 1968, and earlier by Bazell 1954. No such criteria figure in the present characterization. In natural phonology, a sound is a phoneme if and only if not all its occurrences can be derived naturally (i.e. through natural interactions of active natural processes of the language) from other sounds. Thus a language learner will take the nasal vowel of a form [m] to derive from a nonnasal vowel by progressive nasalization from /m/, or regressive nasalization from a (deleted) following nasal, rather than learn to produce the vowel deliberately as nasal. But if the language presents him with forms like [nɑ], which entails the suppression of progressive nasalization; or [en], which entails the suppression of regressive nasalization; or even [en], which entails the suppression of a process deleting a
neutralizing processes apply prior to all non-neutralizing processes. This criterion of natural derivability obviates reference not only to the structuralist criteria but also to the increasingly ad hoc and complex criteria of generative phonology (Zwicky 1972 lists about thirty of these). It furnishes an irreducible account of the acquisition of phonology, in which often a simple form will suffice to show the child that a certain sound is phonemic in the language he is learning; compare the [zeka] example on page 29. (However, if it makes matters simple for the child, it does not do so for the linguist, since to apply the criterion of natural derivability requires the linguist to understand fully the processes and derivations in question.)

At the same time, the criterion seems to me to capture the sort of phonemic representation that the structuralists aimed at but which their criteria failed to characterize adequately. It allows, for example, for desirable non-linear mappings like [kaf] with /kant/ can’t, non-invariant mappings like the flap of [dr] three with /r/ and the flap of [er] Eddy with /d/, and so forth (compare Chomsky 1964). And it quite naturally rules out mappings which the structuralists could rule out only with vague or ad hoc criteria (like the identification of the complementary [b] and [p] of English) since there is no natural process which makes initial /n/ into [h] or final /h/ into [n] (compare Bazell 1954).

16 (page 28) See footnote 14 of the preceding paper.

18 (page 31) I am assuming that, in the absence of auxiliary hypotheses, obstructions represented as unspecified as to voicing would surface, once the process that previously governed their surface voice specification was suppressed, as randomly voiced or voiceless. This assumption seems to surprise adherents of the archisegmental theory, I suspect because they did not consider it to have any empirical consequences.

16 (page 32) This example may be a bad one, because it is difficult to show that all [n]s are naturally derivable from /n/ plus velar in current English, in the face of examples like singer versus lnger, and because tests designed to get at the psychological situation give conflicting results in this case. But it is a clear example, and the facts clearly hold true in earlier English, and they are commonplace in many languages. Or one can substitute the dental [n] of plinth for [n] throughout the discussion.

Some writers on natural phonology have presented it as having all phonemic representation not as a level but as the most "concrete" (superficial) representation of a form which is accessible to the phonetically unsophisticated speaker's awareness. This of course follows Sapir's characterization of the phoneme as a percept,) to the phonetically unsophisticated speaker's awareness. This of course follows Sapir's characterization of the phoneme as a percept.) Then it must follow that allophones can find no place in lexical representation: what is not perceived cannot be recorded in memory.
social, or stylistic variants. In variants of a single speaker, such as those cited in Chapter II for the phrase 'divinity dudge', we find a continuum ranging from the "clearest" to the "most pronounceable" forms. This is typical of variants over syntagmatic processes; variants over paradigmatic processes, such as the following forms of 'head' observed by Donegan and me in the speech of natives of the Great Smoky Mountains, are ranked in the opposite order: 

[\texttt{[\textit{h\textbackslash eg}]}] - [\texttt{[\textit{h\textbackslash eg}]}] - [\texttt{[\textit{h\textbackslash ed}]}] - [\texttt{[\textit{h\textbackslash ad}]}] - [\texttt{[\textit{h\textbackslash ad}]}]. Here the later forms are the "clearest" limited in many other derivatives to expressive speech, though children often use these conservative forms (and this is true both of paradigmatic and paradigmatic derivations) in their everyday speech.

21 (page 47) Further differences between rules and processes may be illustrated by the rule adjusting the pronunciation of the Latinate prefix *im-* (*im-*, *im-*, etc.) versus the process assimilating alveolar stops (including nasals) to the point of articulation of a following stop.

1) The rule is limited to a specific morpheme. The process applies regardless of morphological identities.

2) The rule is obligatory (*impolite, *inrelevant); the process is obligatory only within a stress-measure (*imjump, *g\textbackslash i\textbackslash n\textbackslash c\textbackslash l\textbackslash e\textbackslash t*), and in wider domains is optional and dependent on style (careful *f\textbackslash i\textbackslash m\textbackslash p\textbackslash h* versus casual f\textbackslash j\textbackslash m\textbackslash p\textbackslash h*).

3) The rule applies before any processes, and before the point in speech processing when tongue-slips occur. The process applies afterward. (Cf. page 65.)

4) The rule refers only to phonemes, never to allophones. It does not change /m/ before velars (incompetent) as it does before labials (impossible), because [n] is (or was in Latin) an allophone while /m/ is (and is in f\textbackslash j\textbackslash m\textbackslash p\textbackslash h/). The process may optionally become [n], but it is the process that brings this about, exactly as it optionally changes /n/ to /m/ in Paris. The process refers to features, and operates on completely novel combinations of these (e.g. in borrowed words or phonetic exercises like [s\textbackslash a\textbackslash t\textbackslash g\textbackslash a\textbackslash l\textbackslash a] - [s\textbackslash k\textbackslash a\textbackslash p\textbackslash g\textbackslash o\textbackslash b\textbackslash a\textbackslash l\textbackslash a]) as well as native phonemes.

5) The rule is insensitive to prosodic structure, e.g. to stress placement (emphatic, impous). The process operates on prosodic rather than segmental domains, as in [\texttt{k\textbackslash a\textbackslash n\textbackslash t\textbackslash g\textbackslash e\textbackslash l}] - [\texttt{k\textbackslash a\textbackslash n\textbackslash k\textbackslash g\textbackslash e\textbackslash l}] (*[\texttt{k\textbackslash a\textbackslash n\textbackslash t\textbackslash g\textbackslash e\textbackslash l}]) can't go, or if \texttt{[\textbackslash d\textbackslash g\textbackslash t\textbackslash g\textbackslash e\textbackslash l}] - [\texttt{[\textbackslash d\textbackslash g\textbackslash k\textbackslash g\textbackslash e\textbackslash l}] (*[\texttt{[\textbackslash d\textbackslash g\textbackslash k\textbackslash g\textbackslash e\textbackslash l}]]) shouldn't go, where strings of alveolar stops within a stress-measure are assimilated entirely or not at all, as if the strings constituted a single long alveolar stop segment. (See Donegan and Stampe 1978.)

6) Historically, the rule was borrowed with the morphemes it applies to, and did not extend to new morphemes. The process, a native constraint, affects all borrowed morphecms.

7) Systematic phonemic representations (Chomsky 1964), i.e. representations whose derivations require the agency of rules as well as processes, fall every test I know that might establish their psychological reality. This is in striking contrast to phonemic and morpho-phonemic representations whose derivations require only processes. For example, two words rhyme if they match in phonemic representation from their stressed syllables to the end of the word, e.g. lens : bends, reading : meeting, mix : sixch (if pronounced /s\textbackslash k\textbackslash s\textbackslash /), step : leap (if pronounced /\textbackslash ep\textbackslash /), etc. Morphophonemic identity is not sufficient without phonemic identity: banned : head rhyme only if the latter is not pronounced /\textbackslash h\textbackslash e\textbackslash d\textbackslash /, etc. But morphophonemic identity is preferred: rhymes like reading : needing are preferred over reading : meeting, and are more frequent. Systematic phonemic identity, however, is totally irrelevant: rhymes like line : sign (cf. signal), revision : division (cf. revise, divide), cram : damn (cf. damnation) are not perceived as differing from those with matched systematic phonemic representations. This suggests strongly that rules do not operate on representations in the way that processes do, and casts doubt on the appropriateness of a "process model" for rules.

(The rhyme test, from my unpublished paper on the phrase, 'Yes, Virginia...' (1968), also shows that phonetic identity is immaterial. Words with phonemically identical phonemes rhyme perfectly even if they are pronounced differently, e.g. if mat and cat are pronounced /\textbackslash m\textbackslash a\textbackslash t\textbackslash / and /\textbackslash k\textbackslash a\textbackslash t\textbackslash /). But words without phonemically identical rhymes do not rhyme even if they are pronounced exactly alike, e.g. if mat and pad are both pronounced as /\textbackslash m\textbackslash a\textbackslash t/ in the verse:

\begin{verbatim}
Upon a mat
Upon a pad
A yellow cat
Sereneely sat.
\end{verbatim}

This would follow, of course, if allophonic properties of sounds are not perceived.)

For further discussion of the rule/process distinction, see Donegan and Stampe 1979 (sections 2, 4).

22 (page 55) The unordered sequential iterative hypothesis of process application presented in this dissertation is basically identical to the one I proposed in a weaker version, in notes 1, 12, and 14 of the preceding paper. If Anderson's dissertation of the same year is intended to allow iterative (ABA) application, which is not clear to me then his hypothesis of 'local ordering' was identical to mine except in its preference for counterbleeding over bleeding order (after Kiparsky 1968). For a review of some of the subsequent developments in ordering theory, together with a revised hypothesis of ordering in natural phonology, see Donegan and Stampe 1979 (section 3).

23 (page 57) For what I hope represents some progress in my understanding of the syllable, see Donegan and Stampe 1978. Donegan and I have stuck by the notation of syllables simply by marking their boundaries (·) rather than, as has become fashionable, representing them as trees, because we know of no evidence that their structure cannot universally be deduced from the relative order and sonority of the constituent segments.
One idea in the present work that still seems to me worth pursuing is the conception of syllable-division, re-syllabication, and de-syllabication as a single integrated process.

24 (page 57) It is the prosodic domain of flapping that accounts for the main differences between speakers. Those who flap in the offset of syllables flap only the first t of entity (ent.i.ty): those who flap in the offset of stress-measures (a group of syllables beginning with a stressed syllable and ending before another stressed syllable) flap both t's; and so on, up to those who flap apical stops (at least if they are voiced) even in the onset of stressed syllables, as in the doctor, or today.

25 (page 60) The relative independence of laryngeal, nasal, position, and color features has not been given sufficient attention. But although I do not think this explanation of the "counterbleeding" relation of vowel nasalization and (nasal) flap deletion is wrong, another more general explanation follows from the proposal that processes apply simultaneously (Donegan and Stampe 1979: see note 27 below). In this case [ds.vIr.e.ti] would change in one step to [ds.vIr.e.a.ti], so that deletion could not bleed nasalization.

26 (page 64) On an asegmental or prosodic view of phonological processing (Donegan and Stampe 1978, Stampe 1979), a process nasalizing sonorants—insofar as it does not distinguish among kinds of sonorants—would read the long string of sonorants in borrowing as a single sonorant, and would nasalize the whole string in one step. The length of the string nasalized would depend on the length of the string input and the prosodic domain of the nasalization process. Of the pronunciations cited, the first admissible one in each column corresponds to a syllable, and the last to a stress-measure. The intermediate pronunciations correspond to two syllables, or all the posttonic syllables, or some such domain which I doubt is a natural constituent of the prosodic structure of this word. I suspect that my previous inability to rule out the intermediate pronunciation led me to consider it admissible; it seems obvious in either event that my judgment is not to be trusted here. (Compare also note 21, part 5.)

27 (page 68) The entire discussion of sequential application of processes here is subject to a quite different interpretation. Feeding would result not only from sequential iterative application, but also from simultaneous iterative application, where all processes apply at once, and then all re-apply, and so forth. On this view, a counterfeeding constraint would be not a constraint on the ordering of a pair of processes (that one process may not follow the other), but as a no-iteration constraint (that one may not reapply). This has the interesting consequence that no process could be both fed and counterfed. Note that this is precisely the situation, hitherto unexplained, described in the text with regard to the process raising [t] before nasals (p. 66f) and the process devocing final obstruents in Joan Velten's speech (p. 67); these processes either apply to no derivative representations (excepting, as noted, rule-derived ones), or they apply to all of them. Numerous additional examples are given in Donegan and Stampe 1979: section 3.

Further consequences of the simultaneous iterative interpretation are presented in note 28 after the discussion of feeding and counter-feeding application.

28 (page 69) Since this was written, most of the diachronic examples that seemed to involve a change from bleeding to counter-bleeding application have been discredited (by Koutsoudas et al., Vennemann and others). The inescapable conclusion seems to be that bleeding and counter-bleeding application are not language-specific options. But the theory I presented in the text provides no way of predicting when one or the other will occur, and the available hypotheses seem in one way or another incompatible with natural phonology. For example, Kiparsky's "opacity" hypothesis (1971) presupposes that all processes are learned.

The simultaneous iterative hypothesis presented in note 27 predicts that, in the absence of universal priority constraints, no process could bleed another. As pointed out in note 25, this is often the desired prediction. But Kisseberth and Kenstowicz's numerous examples of bleeding order require some restrictions of priority. We have in natural phonology two universal priority constraints which are well-evidenced: rules before processes, and fortitions before lenitions (note 12). Applied to the simultaneous hypothesis, the latter constraint predicts that a fortition will bleed any lenition it can (see the example of batted in note 12). This most frequently occurs when the fortition separates an assimilating and an assimilated segment, thus "bleeding the context" of the assimilation, as Kisseberth and Kenstowicz put it (p. 68 above).
ADDITIONAL REFERENCES


