

## The Study of Natural Phonology



Patricia Jane Donegan and David Stampe

### 1. INTRODUCTION

Natural phonology is a modern development of the oldest explanatory theory of phonology. Its diverse elements evolved in nineteenth-century studies of phonetics and phonetic change (Sweet, Sievers), dialect variation (Winteler), child speech (Passy, Jespersen), and synchronic alternation (Kruszewski, Baudouin), and developed further, still without integration, in twentieth-century studies of dynamic phonetics (Grammont, Fouché) and phonological perception (Sapir, Jakobson). Its basic thesis is that the living sound patterns of languages, in their development in each individual as well as in their evolution over the centuries, are governed by forces implicit in human vocalization and perception.

In the modern version of the theory (Stampe 1969, 1973a),<sup>1</sup> the implicit phonetic forces are manifested through processes, in the sense of Sapir—mental substitutions which systematically but subconsciously adapt our phonological intentions to our phonetic capacities, and which, conversely, enable us to perceive in others' speech the intentions underlying these superficial phonetic adaptations. The particular phonological system of our native language is the residue of a universal system of processes reflecting all the language-innocent phonetic limitations of the infant. In childhood these processes furnish interim pronunciations which, until we can master the mature pronunciation of our language, enable us to communicate with parents, siblings, and other empathetic addressees. Gradually we constrain those processes which are not also applicable in the mature language. (In multilingual situations, as the languages are sorted out by the child, so are the processes, so that ultimately a different subset of the universal system governs each native language—cf. Major

1977.) From adolescence, usually, there is little further change, and the residual processes have become the limits of our phonological universe, governing our pronunciation and perception even of foreign, invented, and spoonerized words, imposing a 'substratum' accent on languages we subsequently learn, and labeling us as to national, regional, and social origins. If we have failed to constrain any childhood process which others do constrain, then we are said to have implemented a regular phonetic change. This innovation may be imitated, ridiculed, or brought to the attention of a speech therapist; more commonly it is simply not noticed except by strangers. This is because we learn to discount superficial divergences in others, even the drastically altered speech of young children, through processes we have ourselves suppressed; we may even be able to apply them in mimicking others, or spontaneously, in baby-talk to an infant or sweet-talk with a lover.

This is a *natural* theory, in the sense established by Plato in the *Cratylus*, in that it presents language (specifically the phonological aspect of language) as a natural reflection of the needs, capacities, and world of its users, rather than as a merely *conventional* institution. It is a natural theory also in the sense that it is intended to *explain* its subject matter, to show that it follows naturally from the nature of things; it is not a conventional theory, in the sense of the positivist scientific philosophy which has dominated modern linguistics, in that it is not intended to *describe* its subject matter exhaustively and exclusively, i.e., to generate the set of phonologically possible languages.

The subject matter of the theory is also appropriately designated natural phonology in that, as Kruszewski first pointed out in his 1881 treatise on phonological alternations, the phonetically natural aspect of phonology (as in the [s]:[z] alternation of German *Haus:Häuser* 'house:houses')<sup>2</sup> is distinct in its nature, evolution, psychological status, and causality from the phonetically conventional aspects, whether the latter have taken on morphological motivation (as in the [au]:[ɔy], [ɑ:]:[ɛ:], [ɔ]:[œ], [u]:[y:] alternations of *Haus:Häuser*, *Rad:Räder* 'wheel:wheels', *Loch:Löcher* 'hole:holes', *Buch:Bücher* 'book:books') or not (as in the [z]:[r] alternation of *gewesen:war* 'been:was'). The same distinctions were drawn by Sapir, particularly in his explanation (1921: chapter 8) of the evolution of umlaut in Germanic nouns from a phonetic process to grammatical process. Natural phonology properly excludes the topic of unmotivated and morphologically motivated alternations. Although these have often been lumped together with natural alternations in generative phonology, they *should* be excluded from phonology

if it can, in principle, furnish no understanding of them. Of course, such alternations typically stem historically from phonetically motivated alternations, and these *are* in the province of phonological theory, as are the factors whereby the phonetic motivations were lost. The natural subject matter of an explanatory theory includes all and only what the theory can, in principle, explain. In the case of natural phonology this means everything that language owes to the fact that it is *spoken*. This includes far more than it excludes. Most topics which in conventional phonology have been viewed as sources of 'external evidence' (Zwicky 1972b) are in the province of natural phonology as surely as the familiar matter of phonological descriptions.

The study of natural phonology was abandoned early in this century, not because of any serious inadequacies, but because the questions about language that had inspired it were set aside in favor of questions about linguistics—its methodology and its models of description. The goal of explanation which had directed natural phonology, as well as parallel studies of other aspects of language, was rejected as unscientific by Bloomfield and his generation, which concentrated its efforts on analytic methodology. For the generation of Chomsky, which has concentrated instead on formal constraints on linguistic descriptions (grammars), the goal of explanation was simply redefined: an explanatory theory is one which provides, in addition to a description of the set of possible grammars (universal grammar), a procedure for selecting the correct grammar for given data (Chomsky 1965:34). Chomsky's model is adopted in some detail from that of the conventionalist philosophers Goodman (1951) and Quine, according to whom reality is "what is, plus the simplicity of the laws whereby we describe and extrapolate what is" (1953, quoted by Halle 1961:94).

Although Chomsky's program is widely accepted, we doubt whether it can achieve even its descriptive goal, universal grammar. The problem, as Chomsky and Halle admit in *Sound Pattern of English* (1968:4), is distinguishing essential from accidental universals. They illustrate by imagining that after a future war only people of Tasmania survive: any accidental property of their language would then be a linguistic universal.<sup>3</sup> The answer to Chomsky and Halle's question of how to tell what universals are essential is that an *essential* universal is one which we can show to follow necessarily from the *essence* of things—one we can explain. To paraphrase Quine, reality is what is, and what naturally follows from what is. Ultimately we cannot know *what* can be without understanding *why* it can be.

It may be objected that if universal grammar is innate, as Chomsky has proposed (1965), then we would have an explanation of language universals. We do not think, though, that linguists find this satisfying, any more than someone asking why man walks erect would be satisfied by the answer that erect stance is an innate trait of man. We might as well be told that it is God-given. The issue of innateness, despite all the debate it has aroused, is entirely beside the point. What we want to know, whether the trait is innate or whether it is universally acquired, is *why*: the question, like the questions that guided Darwin, is a question of *value*.

Distinctive value was the foundation of the structuralists' functional definitions of the phoneme as an oppositive element (Saussure 1949), definable in terms of its distinctive features (Jakobson 1932a, Bloomfield 1933). This relativistic conception of phonemes, which provided a rationale for concentrating just on the differences capable of distinguishing words, is understandably appealing to the linguist confronted by a growing but somehow irrelevant mass of instrumental phonetic detail. But words are not only distinguished by sounds, they are made up of them. It is no less important that the sounds that constitute words be distinguishable than that they be pronounceable, combinable, and perceivable (articulate, audible). Jakobson (1942) and Martinet (1955) attempted to explain the various centrifugal (polarizing, dissimilative) tendencies in phonology in terms of this distinctiveness principle. But we have shown in our studies of vowel shifts that these tendencies apply to the nondistinctive as well as the distinctive features of sounds, and that they very often end in the merger of phonemic oppositions (Stampe 1972a, Donegan 1973a, 1976). There are perfectly good *phonetic* explanations of centrifugal tendencies, as diachronic phoneticians such as Sievers (1901:282), Fouché (1927:21–24 et passim), and Grammont (1933: 229, 238, 269ff.) had already pointed out. More important, the distinctiveness principle obviously cannot explain the opposite, centripetal tendencies behind assimilation and reduction, which, as Saussure (1949) had emphasized, are destructive of phonological (and secondarily, grammatical) structure.

Thus, for example, in opposition to the polarizing tendency whereby all spirants become stops, there is an assimilative tendency whereby stops become spirants adjacent to open sounds like vowels. We might account for the first tendency as follows: stops are in themselves easier to produce than spirants, which require a more controlled approximation of the articulators; perceptually, stops present a sharper contrast with adjacent vowels. As for the second, the articulation of spirants requires shorter

travel of the articulators between adjacent vowels than that of stops.<sup>4</sup> Both tendencies are real, both are functional, and both are necessary parts of an understanding of phonology. We have to understand not only why a Tamil speaker, for example, hears a spirant as a stop, but also why, between vowels, he pronounces a stop as a spirant. The discrepancy between the sound perceived and intended, and the sound pronounced, is simply phonology.

This tension between clarity and ease is one of the most obvious, and oldest, explanatory principles in phonology. Modern theories, however, to the extent that they incorporate analogous principles, tend to make them monolithic, like the principle of distinctiveness in structuralism or simplicity in generative phonology. This is because they are conceived in modern theories as conventional rather than explanatory principles: they are intended to furnish a choice between alternative descriptions, in accordance with the conventionalist framework we have described. In that framework, positing conflicting criteria would be like pitting Ockham's razor against an anti-Ockham who multiplies entities as fast as the razor can shave them off: it would defeat their purpose of evaluating alternative analyses. But an evaluation criterion, necessarily monolithic, cannot replicate conflicting explanatory principles. The structuralist criterion of distinctiveness predicts that the optimal language should lack contextual neutralizations altogether, and the generative criterion of simplicity predicts that it should lack 'rules' altogether. This is the impasse that confronted Halle (1962) and Kiparsky (1965) in their attempts to furnish a generative explanation of the nature of sound change: the simplicity measure predicts that change would involve the loss of old phonetic substitutions, rather than the accretion of new ones. Postal, who proposed (1968) that rules are added to a grammar for the same reason that manufacturers add fins to cars (presumably for no reason at all) seems at least to have grasped the hopelessness of explaining sound change in terms of the simplicity of grammars.

The basic difficulty is that descriptive models like structural and generative phonology, by the very fact that they provide models for the empirical analysis of languages, provide explanations for what is *learned*. But there is no evidence that the processes which govern phonetically motivated alternation and variation, children's regular sound-substitutions, and phonetic change are learned. On the contrary, there is massive evidence that they are natural responses to phonetic forces, centripetal and centrifugal, implicit in the human capacity for speech production and perception. As Passy (1890), Baudouin (1895), and

many others have observed, the child has many phonetically motivated substitutions, but few, if any, morphologically motivated or unmotivated ('traditional') substitutions; in learning language, he suppresses the inappropriate natural substitutions and acquires the appropriate conventional ones. From this observation it is a small step to the conclusion that phonetic changes must arise from the failure of children to constrain certain natural substitutions, and that variation in adults, another likely source of change, must result from natural substitutions which the individual has suppressed in certain speech styles but which apply inadvertently in other styles.

This account of the correspondences of phonological development, variation, and change explains much that was inexplicable in the structuralist and generative frameworks. For example, according to Jakobson's model of phonological development (1942), the child's phoneme system grows by the step-by-step mastery of oppositions. But there is much evidence not only that the child's mental representations cannot be deduced from his utterances, according to the structuralist definition of the phoneme, but also that they correspond rather closely to adult phonemic representations (Stampe 1969 and forthcoming, Edwards 1973). This means that the child's mapping of phonemes onto phonetic representation, with its massive neutralization of oppositions, is far more complex than the adult mapping. In terms of generative phonology, the child has many more 'rules' than the adult. This paradox disappears when we recognize that the mappings are not rules at all, but simply natural processes motivated by the innate restrictions of the child's phonetic faculty.

An analogous paradox exists in the fact that inattentive (i.e., ordinary) speech presents far more substitutions than attentive speech (Dressler 1972, Stampe 1973a). Variants like [kept ~ kep] *kept*, [prabibli ~ prabli ~ prali ~ prai] *probably*, [ædð'noʊ ~ mmm] *I don't know* pose obvious problems for the structuralist conception of the phoneme, and also for the generative conception of 'rules', since it is when attention is relaxed that 'rules' are multiplied. To avoid these embarrassments, both theories have restricted themselves to artificial phonetic representations (the 'clarity norm' of Hockett 1955, the Kenyon and Knott citations of Chomsky and Halle 1968), dismissing actual speech as 'ellipsis' (Jakobson) or 'performance' (Chomsky and Halle), and thereby failing to account for the main characteristics of the unique 'accents' of the languages under description. The view that speech processing is mediated by systems of natural processes, on the other

hand, predicts that actual speech should normally be quite elliptical and variable. The extent of stylistic and dialectal variability has been brought out quite clearly by the studies of Labov, Bailey, and other 'variationists'.

When loanwords are adapted to the native system, they undergo systematic substitutions, many of which cannot be explained by a system of rules based on native alternations. For example, speakers of many languages which lack final obstruents devoice these when they are pronounced in foreign loanwords. Obviously, neither structuralist phonotactics or generative morpheme-structure constraints would posit, in a vowel-final language, a rule devoicing final obstruents. But devoicing of final obstruents is a natural process, and since it is one which would not be suppressed in the acquisition of a language lacking final obstruents altogether, this devoicing in foreign words is precisely what we should expect. (See further Ohso 1972, Lovins 1973, 1974.)

The summary paradox confronted by the view that all phonological alternations are rule-governed is that the vast majority of such 'rules' are harder to disobey than to obey. The phonetically motivated devoicing of final obstruents as in *Hau[s]:Häu[z]er*, to recall Kruszewski's examples cited earlier, is a 'rule' that is difficult if not impossible for a German speaker to disobey—even, for example, in pronouncing English *cows*. Only phonetically unmotivated rules, like the vowel umlaut of *Haus:Häuser*, which is conditioned by the *-er* plural (contrast the singular noun *Maus* 'molt'), can be disobeyed without phonetic effort. Or consider an English example, the difficulty of suspending the phonetically motivated devoicing of the [z] of *is* to [s] in [ðætso:l] *that's all*, versus the ease of suspending the phonetically unmotivated voicing of the [s] of *house* to the [z] of the plural *houses*. Like umlaut, this voicing of plurals must be learned: we have all heard children say *hou[s]es*, and perhaps occasional adults. But we have not heard anyone, least of all a child, say [ðætz], and no one has reported a child who failed to devoice final obstruents in acquiring German, or Russian, or any other language which devoices final obstruents. In fact, every child we are aware of whose earliest pronunciation of English has been recorded has regularly devoiced final obstruents, e.g. Joan Velten's [nap] *knob*, [bat] *bad*, [ut] *egg*, [duf] *stove*, [wus] *rose* (Velten 1943). And those who continue in adulthood to devoice final obstruents, e.g. [bæ:t], for *bad*, require some effort not to devoice, particularly in situations where the voiced obstruent cannot be released, e.g. in [bæ:tnu:z] *bad news*. There is nothing to indicate that phonetically motivated alternations are governed by rules which are acquired.

The basic forms as well as the alternants of words conform to natural phonetic restrictions. Just as English [z] becomes [s] after a tautosyllabic voiceless segment as in [ðætso:l] *that's*, likewise there are no simple English words or syllables ending in voiceless segment plus [z]; there are [frits] *Fritz* and [faks] *fox*, but not \*[fritz] or \*[fakz]. There are no languages in which all final obstruents are voiced,<sup>5</sup> but many in which they are all voiceless, e.g., Vietnamese. Such restrictions, although they do not result from substitution, can be explained, like alternations, as process-governed. This explanation is strongly suggested because, as is well known, speakers of Vietnamese and other such languages devoice final voiced obstruents in foreign words. For speakers who never encounter such words, the devoicing process remains merely a tacit restriction, but one which limits the universe of words which they might coin. Incidentally, Vietnamese, a near-perfect example of an isolating language, is the sort of language which is sometimes said to have no phonological rules; but one need only listen to the 'accent' a Vietnamese imposes on French or English to see that, rules aside, there are as many processes governing the phonology of speakers of Vietnamese as of other languages.

It was with respect to restrictions like these on basic forms (called phonotactics in structural phonology and morpheme structure in generative phonology) that generative phonological theory, according to Chomsky (1964, 1965), first achieved the level of 'explanatory adequacy'. He argued that the admissibility in English of the nonoccurring word *blick* and the inadmissibility of *bnick* is explained by the theory's evaluation procedure, which rejects any rule that predicts fewer features than are required to state it (viz. 'Liquids are nonlateral after initial voiced labial stops before high front lax vowels before voiceless velar stops') and accepts any rule which predicts more ('Consonants after initial stops are non-nasal'). The first rule is true only of *brick* while the second is true of *beautiful*, *bwana*, *brick*, and for that matter, *blick*; hence, *bnick* is ruled out, but *blick*, should it turn up, is ruled in. Now, most people accept something as an explanation only if there is some reason to believe that this something does in fact obtain. Chomsky gave no reason whatever to believe that such a feature-counting criterion obtains in the language acquisition of children. Nor did he explain how it would follow from this explanation that *blick* is easy to pronounce and *bnick* is not.<sup>6</sup>

The reason, we believe, is that syllable margins are phonetically optimal, *ceteris paribus*, when their constituents present the greatest

mutual contrast of sonority. The contrast of stop is greatest with vowels (stop-vowel syllables are universal), then glides, liquids, nasals, fricatives, and finally, other stops. This is of course the hierarchy of prominence/aperture of Sievers, Jespersen, Saussure, Grammont, et al. After syllable-initial stops English admits everything from vowel through liquid: [ki] *key*, [kju] *cue*, [kwɪt] *quit*, [kru] *crew*, [klu] *clue*, \*[kn . . .] (except in Scots *knife*, etc.), \*[ks . . .], \*[kt . . .] (inadmissible in all dialects). *Blick* falls on the near side of the cutoff point and *bnick* on the far side.<sup>7</sup>

## 2. NATURAL PROCESSES

**2.1 Ontology and teleology.** "Children's speech has far more neophonic alternations . . . than the normal language. As children's language comes to resemble that of adults, the child . . . loses the most innovative variants." (Baudouin 1895:210). Writing for an age which had not grasped the concept of synchronic process, Baudouin spoke of neophonic alternations where we would speak of phonetically motivated processes. Such processes explain not only alternations (e.g., our daughter Elizabeth's [hʌgi ~ hʌk] *hug*, with the word-final obstruent devoiced unless an epenthetic vowel protected it), but also children's non-alternating substitutions as compared with adult speech (e.g., Joan Velten's invariably devoiced [z] in [wus] *rose* (Velten 1943:287)).<sup>8</sup> Baudouin recognized that such alternations in adult languages, as in German [tɑ:gə]:[tɑ:k] 'days:day' were not simply imitated but were developed independently by the child (1895:209); again we would add that this is true of the corresponding restrictions in languages whose final obstruents are invariably voiceless, as in the suffixless Vietnamese above, whose speakers' devoicing of voiced final obstruents in foreign words we have alluded to already. The total system of processes, then, governs both superficial alternations and underlying restrictions.

This dual function can even be performed by a single process, for example the process deleting [h] as in [(h)ɪstɔ:rɪk] *historical* and [(h)weɪl] *whale*. The deletability of [h] is an inverse function of stress and an inverse function of the sonority of the following segment. We will consider only the latter here. In Old English [h] did not precede obstruents. In Middle English it was deleted before other consonants, e.g., from OE [hnutu] *nut*, [hlæxxan] *laugh*, [hrɪŋ] *ring*. Some Modern English speakers (e.g., DS) hold the line here, but others (PJD) delete

[h] before [w] as in *whale*, and still others (HJ) delete it also before [j] as in *hue*. In all these dialects optional deletion can occur in relaxed speech, especially under lighter stress. So we have the following distribution:

	<i>whale</i>	<i>hue</i>	<i>high</i>
DS	[hweɪl ~ weɪl]	[hju ~ ju]	[haɛ ~ aɛ]
PJD	[weɪl]	[hju ~ ju]	[haɛ ~ aɛ]
HJ	[weɪl]	[ju]	[haɛ ~ aɛ]

For DS, *whale:wail*, *hue:you*, and *high:eye* are phonologically distinct, though they may merge phonetically. For PJD, *whale* and *wail* are homophones, and for HJ likewise *hue* and *you*. The [h]-deletion process governs the variation [haɛ ~ aɛ] and at the same time governs the restriction against basic forms like [hweɪl]. PJD and HJ delete [h] from new *wh*-words learned by ear from speakers like DS, and in fact often seem not to perceive the [h] in the first place. We can predict from these facts an English of the future, in which [h]-deletion will apply absolutely, and even *high* and *eye* will be homophonous. This has occurred in the Romance languages, some Indic languages, Greek, and so on, always with the children leading the way. This is not to say that the Italian speaker, for example, is still deleting the [h] of *Romulus* and *Remus*; he does not even confront an [h], except in foreign words. But in these words he deletes it just as we delete the [h] of *high* in relaxed styles of speech. The process applies without premeditation in our ongoing speech, and the hierarchies, within phrases, are perfectly observed: we hear *huge white house* pronounced [hjuɪ hwaɪt haʊs], [hjuɪ waɪt haʊs], [juɪ waɪt haʊs] or even [juɪ waɪt aʊs], but not \*[hjuɪ hwaɪt aʊs], etc.

The fact that processes operate in ongoing speech production is most clearly evidenced by their application to here-and-now, 'non-lexical' outputs of secret-language rules and slips of the tongue (Bond 1969, Fromkin 1971, Stampe 1973a). For example, when the intended words *mostly* [moʊstli] or *mainly* [meɪnli] are unexpectedly replaced by the blend [mōʊnli] or [meɪstli], the nasalization of the exchanged stressed vowel depends not on its ordinary phonetic quality ([ou] vs. [ē]) but on whether or not its novel, slipped context includes a nasal. This suggests that lexical /meɪnli/ or /moʊstli/ become by tongue-slip (specifically, by vowel exchange) /mōʊnli/ and /meɪstli/ and that then nasalization applies—or fails to—producing [mōʊnli] or [meɪstli]. Tongue-twisters—sequences which force tongue-slips—show the same



thing: *sane lad slain* [sɛ̃ɪn læd slɛ̃ɪn], when repeated rapidly, often comes out [seɪd læn . . .]; when the final consonants are switched, the preceding vowels agree with them with respect to nasality even though the vowels themselves keep their original positions. That is, /seɪn læd/ becomes /seɪd læn/ by the slip, and then nasalization produces [seɪd læ̃n]; the inadmissible \*[sɛ̃ɪd læ̃n] never occurs. With our students we have observed hundreds of slips, obtained from tongue-twisters designed to produce sequences phonetically inadmissible in English. In not one case was such a sequence observed to be articulated.

It should not be supposed from this that processes are peripheral, physical events—merely the results of articulatory mistimings or of over- or under-shootings of articulatory targets. We have no reason to suppose that the articulatory musculature or its peripheral innervation can make the kinds of adjustments processes involve (Lashley 1951). Anticipatory substitutions, in particular, suggest that the substitutions occur in the central nervous system—i.e., that they are mental substitutions. The very suppressibility of processes argues for their mental nature—the English-speaking child (Velten 1943) learns not to de-voice final stops after all. And note that processes apply even in silent mental speech, in which purely physical inaccuracies of articulation would play no part (Stampe 1973a).

But although processes are mental substitutions, they are substitutions which respond to physical phonetic difficulties. To illustrate this, it is well to look at variants, like the ordered pairs [tɪn kæn] ~ [tɪŋ kæn] *tin can*, *rai[n,m]bow*, *se[t,p]back*, *re[d,b]man*, etc., which provide direct access to the inputs as well as the outputs of processes, and thus reveal the teleology of substitutions more directly than categorical alternations do. The phonetic motivation of a variable process is usually introspectively quite perceptible: the (more) basic representations (the left-hand members of the paired variants above) seem more difficult to produce than the derivative (right-hand) ones.

**2.2 Natural application of processes.** Processes apply in ways that follow from their nature and teleologies. First, since processes represent responses to phonetic difficulties, it follows that if a certain difficult representation undergoes a substitution, all other representations with the same difficulty will, *ceteris paribus*, undergo the same substitution. This explains why processes operate on 'natural classes' of segments. To this observation, which dates from the Neogrammarians (Sievers 1901:7), we should add that they operate over natural prosodic constituents—

syllables, accent-groups, words, etc. (Donegan and Stampe, forthcoming).

For example, our English reflects the following process:

- (1) Sonorants become nasalized before nasalized segments within a stress-group, but only optionally across syllable boundaries.

Thus we pronounce *rallying* [ræl.i.ɪŋ] or [ræl.ɪ.ɪŋ] or [ræl̃.ɪ.ɪŋ], *relying* [ri.læ̃.ɪŋ] or [ri.læ̃.ɪŋ] but not \*[rɪ.læ̃.ɪŋ] (*re-* being outside the stress-group containing nasality), *rollicking* [rá.lɪk.ɪŋ] but not \*[rá.lɪk.ɪŋ] (*k* being a non-sonorant). Natural classes are not a matter of descriptive simplicity, as suggested by Halle (1962), but a matter of fact: nasalization applies to novel sonorants and before novel nasals, as in our pronunciations of [y] in French *lune* and the vowel before [ɲ] in Spanish *cañon*.

And natural classes cannot be explained as a matter of cognitive simplicity (what structuralists called pattern congruity and generativists call generality) in the acquisition of a 'rule'. The natural classes a process operates on have a natural connection. Nasalization never applies just before non-nasals, or before aspirates, or in alternate syllables. The natural connection is the phonetic teleology of the process.

Each natural process, then, applies to a natural class of representations (namely, all representations which share a common articulatory, perceptual, or prosodic difficulty to a common degree), and each process makes substitutions by altering a single phonetic property to remedy the difficulty. Since the substituted sound should, in each case, be as perceptually similar to the original target as possible, it follows that the changes processes make will be minimal: a process normally changes only one feature. This means that apparent two-feature changes take place in two steps—for example, a change in which [U] → [ʌ] is in fact [U] → [i] → [ʌ] or [U] → [ɔ] → [ʌ].

It has been suggested that such series of simple changes are changed into single substitutions by an operation called 'rule telescoping' (Hyman 1975), so that (consecutive) processes A → B and B → C are collapsed to A → C. This may be true of learned rules, which lack phonetic motivation and which may therefore substitute one phoneme for another regardless of the number of feature changes involved. But processes do not telescope, because distinct processes have distinct phonetic causalities. To establish the telescoping of two processes A → B and B → C would require examples of languages in which A → C while B does not become C.

Stampe (1973a) has cited the opposite case of an American speaker who pronounced syllable-final /l/ always as [ɫ], but who, in attempting to actually pronounce the light [l] of German, frequently said [t̚]; this is precisely what we should expect if, comparing other speakers whose final /l/'s vary between careful [t̚] and careless [ɫ], we assumed that this speaker had the same processes, [l] → [t̚] and [t̚] → [ɫ], as the others, but that in his case the [t̚] → [ɫ] process was obligatory except when overcome in aiming at the totally foreign pronunciation [l].

The principle that each process has its own phonetic motivation (and different motivations mean different processes) explains the mutual dependencies or independencies we find in certain sets of substitutions. Inputs to distinct processes are independently difficult and vary independently: in *what kind*, [hw, w], [t, k], and [d, Ø] are independent variables because each of the substitutions involved (hw → w, t → k/ \_\_\_k, and d → Ø/n \_\_\_#) represents a response to a different phonetic difficulty.

But, on the other hand, inputs to a single process with a single motivation, if not equally difficult (as *ti[n]can*, *rai[n]bow*, etc. seem to be), are unilaterally hierarchic in difficulty and vary dependently: the [h, Ø] variation discussed above in *huge white house* illustrates this dependency, as does its dependence on relative stresslessness; in *hēr hénhòuse*, pronouncing [h] in *hēr* entails pronouncing it in *hòuse*, which in turn entails pronouncing it in *hén*, so that [ø]ér[h]én[ø]duse is admissible but \*[h]ér[ø]én[h]duse is not. Similar (in this case identical) substitutions in a dependent relationship like this are single responses to a single phonetic difficulty present in different degrees. Only in such cases do we say the substitutions result from a single process. We should expect that substitutions which respond to a given difficulty apply to the more-difficult segments if they affect the less-difficult ones, though such substitutions may affect *only* the more-difficult segments in the class. That is, processes are subject to *implicational hierarchies of applicability*. For example, we have argued (Stampe 1972a, Donegan 1973a) that the vocalic features of palatality ("frontness") and sonority ("openness") are articulatorily and acoustically incompatible features, and that there is a process which resolves this difficulty in favor of sonority by substituting non-palatal vowels for palatal ones (a → α, ε → Λ, i → ÿ). Of [a, ε, and i], it is most difficult to maintain a palatal character for the open [a], less difficult for mid [ε], and quite easy for close [i]; and correspondingly, any language in which depalatalization changes [ε] to [Λ] will also change [a] to [α] and any language in which the process changes [i] to [ÿ] will also change [ε] to [Λ] and [a] to [α]. The depalatalization of a lower vowel,

however, implies nothing about the depalatalization of any higher vowel; i.e., the implications are unilateral—and this follows from the fact that the scale of difficulty is unidirectional. Implicational conditions on process applicability are also discussed by Chen (1974b), Donegan (1976), Neeld (1973), Schourup (1973b), Zwicky (1972a) and others.

Each process is sensitive to a number of *different* hierarchical constraints on its application. For example, not only are lower vowels more susceptible to depalatalization than higher ones, but lax vowels are more susceptible than the corresponding tense ones, and labiopalatals are more susceptible than 'pure' palatals. In each case the more susceptible is the less palatal.

But it should be noted here that even though their categorizations are based on physical realities, the phonological features in terms of which processes are specified are mental categories—not just physical scales. For example, labiality in vowels is a feature which corresponds to an articulatory gesture of constricting the lips. Different labial vowels differ in degree of constriction, and there are phonological processes which are sensitive to this difference: less labial vowels are more susceptible to delabialization, less likely to cause assimilative labialization or dissimilative delabialization of adjacent segments, etc. But these processes are not applied as if labiality were a simple physical scale corresponding directly to degree of lip constriction. Instead, processes which depend on degree of labiality depend on height (higher labial vowels are more constricted than the corresponding lower ones), or on tenseness (tense labial vowels are more constricted than the corresponding lax ones), or both. To cite just one example: mid (and presumably tense) [ō] unrounded in one language (IE \*ō > Sanskrit ā) where no high vowels unrounded, but high lax [u] unrounded in another (English [u] → [ʌ]) where no tense vowels unrounded. One might be tempted to hypothesize that IE \*[ō] was less rounded than \*[u], but that English [u] was less rounded than [ō]. But categorical differentiations of this sort in the application of processes occur so often that we are drawn to the conclusion that processes are not dependent on purely physical characteristics, but rather on our mental categorizations of these physical characteristics. If it were otherwise, processes would apply regardless either of their perceptual or of their articulatory consequences, since it is in the mental categorizations of sounds that their double nature is unified.

Speech styles vary, and speech is used with different degrees of attention and emotion. Consequently, different degrees of difficulty—and





or variable one. It is not the form of the process, but its function, that matters.

**2.4 Types of processes.** According to a traditional but well-evidenced typology, there are *three* main types of processes, each with distinct functions:

(a) *Prosodic processes* map words, phrases, and sentences onto prosodic structures, rudimentary patterns of rhythm and intonation. Insofar as syllabicity, stress, length, tone, and phrasing are not given in the linguistic matter, they are determined by the prosodic mapping, which may most easily be described as an operation in real-time speech processing of which setting sentences to verse or music are special cases. (Stampe 1973b, and compare Goldsmith's paper and references, in this volume.) The application of prosodic processes is the most important factor in the living phonological pattern of a language and its long-range phonological 'drift'; the selection of segmental processes is largely determined, even in childhood, by the way segmental representations are mapped onto prosodic structure in speech (Major 1977, Stampe and Donegan forthcoming). However, since the remainder of our discussion is mostly concerned with segmental issues, we must turn to the processes which govern segments.

(b) *Fortition processes* (also called centrifugal, strengthening, paradigmatic) intensify the salient features of individual segments and/or their contrast with adjacent segments. They invariably have a perceptual teleology, but often incidentally make the segments they affect more pronounceable as well as more perceptible. Dissimilations, diphthongizations, syllabifications, and epentheses are fortition processes. Some fortition processes may apply regardless of context, but they are particularly favored in 'strong' positions, applying especially to vowels in syllable peaks and consonants in syllable onsets, and to segments in positions of prosodic prominence and duration. Similarly, they apply in situations and styles where perceptibility is highly valued: attentive, formal, expressive, and *lento* speech.

(c) *Lenition processes* (also called centripetal, weakening, syntagmatic) have an exclusively articulatory teleology, making segments and sequences of segments easier to pronounce by decreasing the articulatory "distance" between features of the segment itself or its adjacent segments. Assimilations, monophthongizations, desyllabifications, reductions, and deletions are lenition processes. Lenition processes tend to be context-sensitive and/or prosody-sensitive, applying especially in 'weak'

positions, e.g., to consonants in 'blocked' and syllable-final positions, to short segments, unstressed vowels, etc. They apply most widely in styles and situations which do not demand clarity (inattentive, intimate, and 'inner' speech) or which make unusual demands on articulation (e.g. rapid tempos).

The fortition/lenition distinction, under various names, is a traditional one in diachronic phonetics. Due to its teleological character it has played no systematic role in modern phonology. But it is indispensable in any attempt at explanation, because almost every phonological process has a corresponding process with exactly opposite effects. For example:

- (2) (f) After nasals, before spirants, a stop is inserted homorganic to the nasal and of the same voicing as the spirant, e.g., [sen(t)s] *sense*, [bæn(d)z] *bans*.
  - (1) Stops after homorganic nasals before spirants (etc.) are deleted, e.g. [sen(t)s] *cents*, [bæn(d)z] *bands*.
- (3) (f) Pretonic resonants are syllabified, e.g. [preɪd] *prayed* → [preɪd] (emphatic).
  - (1) Pretonic syllables are desyllabified, e.g., [preɪd] *parade* → [preɪd] (casual).
- (4) (f) Achromatic syllabics assume a color (palatal/labial) opposite that of their off-glides, e.g., [æɪ] *I* → [ɒɪ] (Cockney), [ʌɪ] *oh* → [eɪ] (affected RP British, occasional U.S.).
  - (1) Achromatic syllabics assume the same color as their off-glides, e.g., [æɪ] *I* → [æɪ], [ʌɪ] *oh* → [o].

Here we have (2) insertion/deletion, (3) syllabification/desyllabification, (4) dissimilation/assimilation in identical contexts, but in each case the fortition (f) typically accompanies strong articulation and the lenition (1) weak. The vowel substitutions in (4) typically accompany longer (f) versus shorter (1) pronunciations, as in [ge:ɪ] *go*, [gʌ:ɪz] *goes*, [gouɪ] *going*. The causalities of the (f) and (1) processes are opposite, reflecting respectively the clarity versus ease principles of traditional phonology.

**2.5 Processes and rules.** It is not the case that all phonological alternations are governed by natural phonological processes. The principles which underlie alternations which are not process-governed—like 'Velar Softening,' 'Tri-syllabic Laxing,' etc.—we refer to as phonological rules. The real nature of such rules is not entirely clear to us, but it is clear

that they differ from natural processes in many important respects.

First, and most importantly, processes have synchronic phonetic motivation and represent real limitations on speakers' productions. Rules lack current phonetic motivation; they are sometimes the historical result of 'fossilized' or conventionalized processes which have lost such motivation (cf. Baudouin's paleophonetic alternations [1895])—German umlaut is an example. On the other hand, processes *lack* positive semantic or grammatical functions, which some rules (like umlaut) do have. (Processes may of course have the negative effect of neutralizing semantically relevant phonological distinctions (*latter/ladder*).)

Processes are 'innate' in the sense that they are natural responses to innate limitations or difficulties; we pronounce *profound* [profãõnd] rather than \*[profãõnd] because we *can't* say the latter without acquiring greater velar precision. But since there is no phonetic reason for saying [profãnditi] instead of \*[profãõnditi], we must say the former simply by convention—because that's what other speakers say: rules are learned.

Processes apply involuntarily and unconsciously, and are brought to one's consciousness only negatively, by confrontation with pronunciations which do not conform to the process, as in second language acquisition. Even the causes of a process may be quite unavailable to consciousness; they may consist in allophonic differences the speaker is quite unaware of. Rules, although they may become habitual and therefore involuntary and unconscious in their application, are formed through the observation of linguistic differences of which the speaker is or was necessarily conscious.

Processes not only govern alternations—they represent constraints on our pronunciations and can be violated only if the speaker makes a special effort (and sometimes not even then). Rules only govern alternations; they often tolerate phonetic exceptions: pronunciations like [æftn] *Afton* and [bõstn] *Boston* violate the rule which deletes the /t/'s of *soften*, *fasten*, etc., but the alternation is nevertheless quite regular for the morphological construction to which it does apply.

Processes apply to tongue-slips, as noted above (2.1), to Pig Latins, to foreign words, etc. (Stampe 1973a). Rules do not ordinarily apply in these cases. This leaves open the question of whether rules in fact *apply* in speech production.

Processes can't be borrowed, any more than speech impediments can be borrowed. If a process in the loaning language produces frequent

alternations in vocabulary which is borrowed into another language, and if certain morphological conditions are satisfied, the borrowers may formulate a rule corresponding roughly to the process in the loaning language. But it is merely a rule, and has none of the properties of a process except those of superficial resemblance (cf. Lovins 1974).

Processes may be optional (variable) or obligatory. Rules, on the other hand, seem always to be obligatory. Apparently, the entirely conventional nature of rules exempts them from the phonetic pressures (toward ease or clarity) of style and tempo variation. Since a rule's application has no phonetic value, rules are no more or less likely to apply in any given style, no matter what its phonetic demands.

Constraints on admissible forms in a language are either phonetically motivated or not. Forms which violate phonetic constraints, e.g., [bnrk], [bax] *Bach*, [vẽ] *vin*, [dhobi], [spõta] etc., are typically adjusted to correspond to these constraints by the phonological processes of the language. Forms which are inadmissible for other reasons—typically accidental or historical—although they are rarely chosen when speakers coin new terms, are adopted without change: e.g., [bwrk], *Houck*, *Cowper*, *Sharnk*, etc. This is because there are no processes active in the language to provide substitute pronunciations in the latter case; other forms the speaker has learned have taught him to pronounce these. (As is suggested by our explanation of [bnrk] versus [blik], being able to say [brk] entails being able to say [bwrk], [r] being less sonorous than [w].)

### 3. DERIVATIONS<sup>9</sup>

In this section we turn to the interactions of processes with each other and with rules. (We have nothing to say here about interactions of rules with each other.) For purposes of discussion we use the taxonomy of interactions of Kiparsky (1968b) of *feeding/counterfeeding* and *bleeding/counterbleeding*, which is presented in many other works.<sup>10</sup>

**3.1 Feeding and counterfeeding.** Because of their specificity, when processes repair one sort of unpronounceability, they sometimes create another. In such cases other processes in turn repair these secondary unpronounceabilities, until a pronounceable representation is obtained. For example, we have the processes

- (5) Elision of nasals before homorganic (tautosyllabic) (voiceless) consonants, e.g. [ment] *meant* → [mēt]—also with regressive nasalization.<sup>11</sup>
- (6) Flapping of intervocalic syllable-final apical stops, e.g. [ðætæp] *that apple* → [ðæɾæp], [bætɪd] *batted* → [bæɾɪd].
- (7) Progressive nasalization of (tautosyllabic) sonorants in unstressed syllables after nasalized segments, e.g., [sɪgnɪl] *signal* → [sɪgnɪ̃l].

Each of these processes 'feeds' the next, in turn, in the processing of a phrase like [plæntɪt] *plant it* → [pl̩æntɪt] → [pl̩æɾɪt] → [pl̩æ̃ɾɪt]. If it were otherwise, the functions of flapping and nasalization would be realized only on basic representations and not on derivative ones.

The hypothetical intermediate steps, which are necessary to explain the pronunciation given,<sup>12</sup> also occur as variant pronunciations in their own right: Zwicky (1972a) cites [pl̩æntɪt ~ pl̩æ̃ɾɪt], Stampe (1973a) [pl̩æ̃ɾɪt] ~ [pl̩æ̃ɾɪt]. Even in speakers who invariably say [pl̩æ̃ɾɪt], intermediate representations can be brought to light in speech situations which block any of the processes; e.g., in secret languages like Ob or Alfalfa which infix [áb] or [ælf] before each syllabic, *plant it* is [plabɪntàbɪt]; in singing, it is often [pl̩æ̃.ɾɪt]. The sequentiality of substitutions like this is confirmed by the fact that no process affects derivative representations unless the process which would create them actually applies; when nasal elision is not applied, or is blocked as in Ob, flapping never applies: (\*[pl̩æ̃ɾɪt]).

But there exist speakers who, though they regularly flap basically intervocalic [t] as in *pat it*, do not flap the derivatively intervocalic [t] of [pl̩æ̃ɾɪt] *plant it*. They apply flapping but constrain it not to apply in sequence after nasal elision. (This is called 'counter-feeding' application: flapping is counterfed by nasal elision.) Such non-universal constraints are relatively common, though Koutsoudas et al. (1971), Vennemann (1974c), Hooper (1976) and others, on various *a priori* grounds, have denied their existence.<sup>13</sup> Although no doubt some examples in the literature actually can be predicted on universal principles (as in 3.2), and others are of dubious synchronic status, many examples, like the present one, seem unavoidable. This seems all the clearer from the existence of single speakers, like ourselves, with sequenced application ([pl̩æ̃ɾɪt]) in informal styles varying with nonsequenced ([pl̩æ̃ɾɪt]) in formal styles,

while basically intervocalic [t]'s are almost invariably flapped: *pat it* is [pæɾɪt], not \*[pætɪt]. Additional examples (cf. also Bailey 1974) will appear below.

We speak of counterfeeding as a *constraint* because, even for speakers with stylistic variation, representations derived from counterfeeding application, like [pl̩æ̃ɾɪt], are more difficult to say than those from feeding, like [pl̩æ̃ɾɪt]. The difference is not so marked as between [pætɪt] and [pæɾɪt] *pat it*, but it is a difference of the same sort—a difference in phonetic difficulty. If flapping of intervocalic [t] has a phonetic motivation (and this can scarcely be doubted), then so does its extension to derivatively intervocalic [t]'s.

Kiparsky (1971) also noted the exceptional character of pronunciations derived by counterfeeding and hypothesized that this makes counterfed processes more difficult to discover (more 'opaque') than fed processes, and so explains the diachronic tendency for counterfeeding to be replaced by feeding. This hypothesis makes sense only if, as is assumed in conventional theories, processes in fact must be discovered. But even if there were any evidence to support that view, we cannot see how the difficulty one might have encountered, as a child, in learning (for example) when to flap could explain *synchronic* variation between counterfeeding in formal speech and feeding in informal, or how it could explain why representations derived by counterfeeding are invariably harder to *pronounce* than those derived by feeding.

As Stampe had already argued (1968b, 1969, cf. 1973a), all these facts, synchronic as well as developmental and diachronic, are accommodated by the theory that the unconstrained application of processes, singly or in concert, is *phonetically* motivated. It is not the processes, but constraints on the processes, which must be acquired. The constraints, however, are equally well motivated, in that they all bring speech closer to its phonological intention. Suppressing the application of a process to the output of another, e.g., not applying flapping to the output of nasal elision (saying [pl̩æ̃ɾɪt] instead of [pl̩æ̃ɾɪt] for *plant it*), like suppressing its application altogether, e.g., not applying flapping at all (saying [pætɪt] instead of [pæɾɪt] for *pat it*), lets this much of the phonological intention (the [t], in these examples) manifest itself in actual speech. Constraints of either sort typically prevent the merger of phonologically distinct representations (e.g. of *plant it* with *plan it* [pl̩æ̃ɾɪt], *pat it* with *pad it* [pæɾɪt]). Of course, this clarity that phonological constraints afford is achieved only by an expenditure of phonetic effort, and it may be sacri-

ficed in less formal styles and less conservative dialects. In formal situations, and for conservative speakers, however, the unconstrained productions may be perceived as careless, inarticulate, or uneducated.

The recognition that constrained as well as unconstrained application of processes is motivated (cf. Donegan 1973b, Kaye 1974, Chen 1972, Kisseberth 1976) was slow to come because it was assumed that constrained applications are due to historical accident, and that only the *relaxations* of constraints (generalization, change to feeding application) are motivated. In particular, it was assumed that processes (or rules) are applied in an order which, unless 'restructuring' occurred, reflected the historical chronology of the processes as sound changes (Halle 1962, Kiparsky 1968b, King 1969). As Bloomfield, who also experimented with ordered-process descriptions (1939), noted, this assumption has been implicit in the methodology of internal reconstruction since the Neo-Grammarians developed it. If changes of the synchronic order of processes are in the direction counterfeeding to feeding, then, by the ordering theory, any counterfeeding application must reflect the original chronology, i.e., an older process counterfed by a younger feeding process.

However, there are counterfeeding derivations in which the counterfed process is the younger, where (to continue using the terms of the ordering theory) rather than going to the end of the line of processes, the newcomer has slipped into line ahead of an older feeding process. In her LSA paper "Southern Discomfort" (1974), Donegan cited recordings of several generations of speakers of the Great Smoky Mountains. Younger speakers have processes, usually variable (and thus unquestionably synchronic), assimilating a mid glide to its syllabic, so that [aɔ] → [a:], as in *house* [haʊs ~ ha:s], *plow*, etc., and diphthongizing [ɔ:] to [aɔ] (via [ɔɔ]), as in *saw* [sɔ: ~ sɔɔ ~ saɔ], *dog*, etc. These are always applied in counterfeeding order: even in six-year-olds the diphthongized [aɔ] is never re-monophthongized by glide-assimilation to [a:]: *saw* is never \*[sa:], *dog* never \*[da:g], etc. In the ordering theory, this situation could only arise if the counterfed process ([ɔ:] → [aɔ]) had entered the language before the feeding process ([aɔ] → [a:]). But records of older speakers indicate that the chronological order was the opposite: most older speakers diphthongized [ɔ:], but monophthongization of [aɔ] is rare in older speech.

Let us cite a more accessible example. English speakers from an ancient date have had the process:

- (8) Apical stops become homorganic to following (tautosyllabic)

stops, e.g. [hændpɪkt] *hand-picked* → \*[hænbpɪkt] → [hæmbpɪkt], and note the inadmissibility of words like \*[hænb].

The following younger process also applies in many dialects:

- (9) Palatal syllabics [æ, ɛ, ɪ] become resp. [æɛ, ɛɪ, ɪɪ] before tautosyllabic [ʃ, ɲ, ɡ], e.g. [bæŋ] *bang* → [bæɛŋ], [fɪʃ] *fish* → [fɪɪʃ], [lɛɡ] *leg* → [lɛɪɡ]. (The generality of the input and context varies by dialect.)

Now, given their chronology, (8) should feed (9), so that, e.g., [mæŋkæŋd] *mankind*, if pronounced with stop assimilation ([mæŋkæŋd]), should also undergo vowel assimilation to [mæɛŋkæŋd]. Some speakers *do* use such feeding-derived pronunciations, at least in informal styles; but most speakers, even those who find it difficult to pronounce *bang* without the transitional glide, feel that [mæɛŋkæŋd] sounds careless and say [mæŋkæŋd] instead. Examples like these show that phonological conservatism can motivate counterfeeding constraints. Since we have argued that phonetic limitations motivate feeding, we must conclude that the synchronic interactions of processes have nothing whatever to do with their history.

Furthermore, we know of no evidence supporting the related assumption of a phonology as a linearly ordered list of processes (or rules) such that a given phonological phrase undergoes each applicable process in turn, and no process applies more than once (Halle 1962, Chomsky and Halle 1968). Sequenced substitutions can be explained simply as the effect of processes applied wherever they are phonetically motivated, without recourse to ordering (Stampe 1969, 1973a, S. Anderson 1969, 1974, Koutsoudas et al. 1971). Chomsky and Halle themselves (1968) cited examples—though they did not grasp their significance—which demonstrate clearly that processes apply more than once, namely processes which feed themselves. Some examples have appeared in our discussion, e.g. sonorant nasalization, both progressive (7) and regressive (1), and apical stop assimilation (8). The device Chomsky and Halle proposed to account for such examples (applying the process to strings of susceptible segments simultaneously) disregards the standard arguments for recognizing sequenced substitutions, and it cannot cope with the contingencies of optional application. For example, regressive sonorant nasalization is obligatory only within syllables, and we there-

fore have such complex patterns of pronounceability (unstarred) in derivatives as \*[kær.i.lɪn] *Carolyn* → \*[kær.i.lɪn] → [kær.i.lɪn] → [kær.i.lɪn] → \*[kær.i.lɪn] → [kær.i.lɪn]. On the iterative interpretation of processes, such examples are precisely accounted for by the simple statement 'nasalize sonorants before (tautosyllabic) nasals.' (For detailed discussion cf. Stampe op. cit., S. Anderson 1969, 1974, Dell 1970, Morin and Friedman 1971, Kenstowicz and Kisseberth 1973, among others.) Furthermore, Stampe (1969, 1973a), S. Anderson (1969, 1974), and Newton (1971) have presented examples of processes applying nonconsecutively more than once in a single phrase (AXA) and applying in different sequences in different phrases (AB/BA).

Most of those who take feeding to be unordered application seem to take counterfeeding (if they accept its existence) to be an ordering constraint on *pairs* of processes, of the form 'B may not succeed A', where B is the (counter)fed and A the feeding process. This presupposes that processes must apply in sequence. But this is not the only possible explanation of feeding. *If processes can apply and re-apply iteratively, then feeding must result even if all processes apply simultaneously.* This would mean that distinct processes apply in the same way as subprocesses of a single process, which are generally accepted to apply simultaneously.

Counterfeeding, on this view, might be a constraint against iterating the counterfed process. Such a constraint on flapping (6), for example, would allow it to apply to basic representations like [pætɪt] *pat it*, but not to derivative representations like [pɛ̃ætɪt] from [plæntɪt] *plant it*. Interpreting counterfeeding as a constraint on iteration differs empirically from interpreting it as a constraint on order, whether linear or 'local' (pairwise), in that it predicts that if a process is counterfed with regard to one feeding process it will be counterfed with regard to all others. For example, if one uses the pronunciation [pɛ̃ætɪt], with flapping inapplicable to intervocalic [t] derived by nasal-elision, then one should also use the pronunciation [fɛ̃ʊtɪt] for [feltɪt] *felt it*, with intervocalic [t] derived by lateral-vocalization, rather than the pronunciation [fɛ̃ʊrɪt]; if [t] and [r] vary in *plant it*, they should vary in *felt it*; and if only [r] occurs in *plant it*, only [r] should occur in *felt it*. Likewise, the vowel assimilation process (9), if it is counterfed by stop assimilation (8), giving [mæŋkæŋd] rather than [mæŋkæŋd] for *mankind*, is also counterfed by the palatalization process in [pæsju] *pass you* → [pæʃju] (not \*[pæʃju]); but speakers who do use the 'fed' pronunciation [pæʃju] also say [mæŋkæŋd]. Stampe (1973a:ch. 2) cited some cases of this sort; for example, speakers (like Stampe) who unexceptionably raise basically prenasal [ɛ]

to [ɪ], as in [pɪn] *pen* (homophonous with *pin*), [jɪm] *gem* (homophonous with *gym*), but do not raise [ɛ] in e.g., [lemɪ] *lemme* (from [letmɪ] *let me* via [leʔmɪ]), where it is derivatively pre-nasal via glottal-deletion, also never raise [ɛ] which is derivatively pre-nasal via regressive nasalization (e.g. [semntɪ], from [sevntɪ] *seventy*) or via vowel and flap elision ([lemaɔt], from [leimaɔt], from [lerimaɔt], from [letimaɔt] *let 'em out*).

Ordering theories, whether linear (e.g. Kiparsky op. cit.) or local (Anderson op. cit.), not only do not explain such cases, they do not even envision them. Consequently the question of whether a process can be simultaneously fed by one process and counterfed by another has not, to our knowledge, previously been raised.<sup>14</sup> We have been unable to find a single such example. Instead, we find many examples, of which the above are a small sample, where, whether a process is fed, or counterfed, or variably fed ~ counterfed, it has this relation to all other relevant processes.

The most straightforward explanation seems to be that feeding, counterfeeding, and variable feeding represent iterative, noniterative, and variably iterative application of single processes, under the hypothesis that all processes apply simultaneously with all others. A noniteration constraint on a process would prevent it from applying to the output of any other process. Such a constraint would pertain to a particular process, without specific reference to other processes, just like any other kind of constraint.

**3.2 Precedence.** We turn now to further relations of processes, the first of which is the 'bleeding'/'counterbleeding' contrast. Process B is 'bled' by process A if there are representations to which both A and B are applicable but A's application changes these so that B is not applicable. This can occur only if the application of A precedes that of B. If B precedes A, or if A and B apply simultaneously, then B is not bled (is 'counterbled') by A, and both apply. Kiparsky (1968b) argued that counterbleeding is natural, historically replacing bleeding. So also Anderson (1974), who follows Kiparsky's generalization (abandoned by Kiparsky since 1971) that maximal application (hence feeding and counterbleeding) is natural. Stampe (1973a) pointed out, however, that maximal application is not self-explanatory, and that if it is understood as phonetically motivated, it would only explain feeding, not counterbleeding: whereas counterfeeding fails to eliminate derivative unpro-



nounceabilities (3.1), neither bleeding nor counterbleeding fails in this regard.

Koutsoudas et al. (1971) had argued, in fact, that bleeding applications do not occur at all, and had proposed that processes apply simultaneously, where possible. Their proposals are made in the framework of a theory that the 'order of application' of processes is universally determined (cf. Koutsoudas 1976 and references). Vennemann (1974c) and Hooper (1976) and some others have assumed variants of this position. We have argued that this is clearly wrong in the case of feeding/counterfeeding examples like those in 3.1. Koutsoudas et al. presented convincing reanalyses of most of the putative bleeding applications in Kiparsky's examples of bleeding-to-counterbleeding changes. To account for other cases they proposed various universal precedence principles, such as Proper Inclusion Precedence (a process whose input is properly included in the input of another precedes the other), Obligatory Precedence (an obligatory process precedes an optional one, cf. Ringen 1972), etc., with various seniorities. We lack space for examples, and will instead give a counterexample:

- (10) (a) Pretonic sonorants are optionally syllabified, e.g., [prei] *pray* → [prei] or [pirei] *pr-ay!* (3f).  
 (b) [r] obligatorily becomes a flap [ɾ] after tautosyllabic [θ], e.g. [θri] *three* → [θɾi].

Here (b) should precede (a) on grounds both of Proper Inclusion and Obligatory precedence, but in fact (a) bleeds (b): [θri] *three* → [θiri] *thr-ee!*, not \*[θiri]. It is probably impossible to refute any precedence hypothesis: when we propose an alternative to account for (10), as we will do in (3.2.1), all we have done is to 'bump' the others into positions of lesser seniority. It is the explanatory value of the hypotheses, not merely their empirical, predictive value, which matters. But Koutsoudas et al. have argued exclusively from empirical grounds.<sup>15</sup>

Kenstowicz and Kisseberth (1971) presented many examples of bleeding interactions, most of them resembling (10), and on the strength of these Kiparsky (1971) revised his original position that counterbleeding is the natural interaction, proposing instead that counterbleeding is opaque: for example, in the counterbled pronunciation of *three*, \*[θiri], the conditions under which [r] is flapped are obscured by the intrusive [i]. As argued in 3.1, this is no explanation if the counterbled process is not learned, and [r]-flapping after [θ] is certainly not a rule.

If processes apply simultaneously, the result is counterbleeding. Examples are legion:

- (11) Regressive nasalization (1) is not bled by nasal-elision (5): [kænt] *can't* → [kæ̃t], not \*[kæt].  
 (12) [t]-flapping (6) is not bled by desyllabification of a following syllabic: [ʃætɾɪŋ] *shattering* → [ʃæɾɪŋ], not \*[ʃætɾɪŋ]; contrast [pætrɪk] *Patrick*.  
 (13) Nor is flapping bled by syncope of a following syllabic: [pʊtɪtɪweɪ] *put it away* → [pʊɾtɪweɪ] (→ [pʊdtɪweɪ]), not \*[pʊttɪweɪ].  
 (14) Pre-nasal [ɛ]-raising (3.1) is not bled by nasal-elision (5), e.g. [sɪ̃t] *sent* (cf. [sɪnd] *send*), not \*[sɛ̃t].

It is significant that the bleeding pronunciations do not seem even remotely possible (learnable) as variants—[ʃætɾɪŋ] for *shattering*, for example, seems pronounceable only in a style or dialect in which flapping does not apply at all. This is in striking contrast to the feeding relations of 3.1, where counterfeeding variants do seem possible even to those who do not use them. And whereas feeding/counterfeeding variation is commonplace in synchronic, diachronic, and developmental phonology, we are not aware of a single good example of counterbleeding/bleeding variation. (Lee [1975–76] has independently observed the non-occurrence of bleeding/counterbleeding variation.) In this respect, therefore, we agree with Koutsoudas et al. (1971) that there is no possibility of extrinsic constraints on processes which are counterbled or bled. Bleeding occurs, we think, only as an incidental result of universal precedence principles.<sup>16</sup> We turn to two of these which are independently evidenced by other relations besides bleeding.

**3.2.1 Fortitions first, lenitions last.** The application of all fortition processes precedes that of all lenition processes. In discussing the fortition/lenition distinction (2.4), we noted that these processes tend to apply in complementary styles. However, many derivations conjoin fortitions and lenitions.

- (15) Nasal-elision (5) occurs only before homorganic consonants, e.g., [læmp] → [læ̃p] *lamp*, [tɛnθ] → [tɛ̃θ] *tenth*, but [wɔ̃rmθ] *warmth* does not become [wɔ̃r̃θ]. However, when a stop is inserted (2f), the elision then occurs: [wɔ̃rmθ] → [wɔ̃r̃mpθ] → [wɔ̃r̃pθ].

- (16) Various dialects have variable fortitions [Uu] → [iu] → [iʉ] → [iɥ], e.g. [tuu] ~ [tiu] *two*, and/or the variable lenitions [iʉ] → [iy] (or [yʉ]) → [yɥ], e.g. [viu] ~ [vyɥ] *view*. Where these co-occur, the processes link to give [tyɥ] *two*, rhyming with [vyɥ] *view* (Donegan 1978).
- (17) The fortitions of (16), or their counterparts in [noʉ] ~ [neɥ] *no*, [nao] ~ [naɔ] *now*, feed the assimilative palatalization of velars, as in [kuʉ] ~ [ciʉ] *cool*, [goʉ] ~ [ʒeɥ] *go*, [kaɔ] ~ [caɔ] *cow*.

In these examples fortitions feed lenitions. We have observed no styles or dialects, however, in which the lenitions are counterfed, so that, for example, [læp] *lamp* co-occurs with [wɔ̄ɾmpθ] *warmth* (15), [tuu] ~ [tiu] *two* with [viu] ~ [vyɥ] *view* (16), or [ciʉ] *cue* with [kiʉ] *coo* (17), and, unlike the counterfeeding pronunciations cited in 3.1, these strike us as not just difficult but impossible to master.

There are many instances of fortitions counterfed by lenitions, for example:

- (18) Casual-speech syncope, e.g., [sɪnɪstɹ] ~ [sɪnstɹ] *sinister*, [tɪmɪθi] ~ [tɪmθi] *Timothy*, never feeds the stop-insertion of (2f): \*[sɪnstɹ], contrast [spɪn(t)stɹ] *spinster*; \*[tɪmpθi], contrast [ʒɪm(p)sn] *Jimson*.
- (19) Tensing of palatal or labial vowels in hiatus, e.g., [i] (not [ɪ]) in *various*, *reality*, *idea*, [u] (not [ʊ]) in *graduate*, *duet*, *suet*, etc., does not apply to vowels put into hiatus by a lenition process, e.g., the allegro deletion of flaps: [dɪrɪt] *did it* ~ [dɪt] not \*[dɪt], [wʊrɪt] *would it* ~ [wʊt] not \*[wʊt].

Normally, feeding application would be more natural than counterfeeding, but pronunciations like \*[tɪmpθi] *Timothy* and \*[wʊt] *would it*, with lenitions feeding fortitions, seem absolutely unnatural.

The following examples, contrasting fortitions and lenitions with identical inputs (20) or outputs (21), are particularly instructive:

- (20) In most dialects chromatic vowels before tautosyllabic [r] are laxed, but in some Middle Atlantic dialects (like PJD's native Baltimore) they are tensed: *beer* [bɪr] vs. [bɪr], *bare* [bɛr] vs. [bɛr], *bore* [bɔr] vs. [bɔr], etc. Various lenitions feed the laxing, e.g. [sɪr] ~ [sɪr] *seer*, [lɛr] ~ [lɛr] *later*, [lɔnmɔr] ~ [lɔnmɔr] *lawn mower*, but lenitions never feed the tensing: [bɪrɹ] ~ [bɪr] *bitter*, not \*[bɪr]; [bɛrɹ] ~ [bɛr] *better*, not \*[bɛr].

- (21) The essentially context-free diphthongization of [æ] to [æɛ] in the South (referred to in 2.3) for many speakers feeds a dissimilation to [aɛ] and even [aɛ], e.g. [baɛd] *bad*, [graɛs] *grass*. But assimilative diphthongization of [æ] to [æɛ] (9) never feeds such dissimilations: [bæɛŋ] *bang*, not \*[baɛŋ]; [hæɛs] *hash*, not \*[haɛs].<sup>17</sup>

The principle that fortitions precede lenitions explains all these otherwise aberrant examples.

Many examples of bleeding, which the simultaneous application hypothesis cannot explain, involve a fortition bleeding a lenition. Example (10) falls under the fortition-first principle, since the optional syllabication of sonorants is a fortition and [r]-flapping is a lenition (assimilation to [θ]).

- (22) In dialects with palatalization of [t,d] by tautosyllabic [r], e.g. [tru] ~ [čru] *true*, [dɾʌŋk] ~ [ʒɾʌŋk] *drunk*, this process is bled by sonorant syllabification: [tɾu] *tr-ue!*, not \*[čɾu]; [dɾʌŋk] *dr-unk!*, not \*[ʒɾʌŋk].<sup>18</sup>
- (23) The English [z] and [d] suffixes, as in [hʌgz] *hugs*, [hʌgd] *hugged*, are devoiced after voiceless segments, e.g. [dʌks] *ducks*, [dʌkt] *ducked*. This assimilation is bled by vowel epenthesis after sibilants, e.g., [kɪsɪz] *kisses* (not \*[kɪsɪs]), and dentals [nɪtɪd] *knitted* (not \*[nɪtɪt]), respectively.<sup>19</sup>

In fact, most of the bleeding examples cited by Kenstowicz and Kisseberth (1971) are of this sort, epenthesis bleeding assimilations. They consider the hypothesis that processes altering syllable structure precede processes dependent on syllable structure, but reject it because, for example, the deletion of a vowel or glide normally fails to bleed the assimilation of a consonant to that vowel or glide, as in Japanese [matɕ(w)] 'wait (for)', [matɕ(i)kamaɛɾu] 'be on the watch (for)', from [mat-] 'to wait (for)'. Since insertions are fortitions, our hypothesis predicts that the former must apply before assimilations.

Deletions are lenitions, however, and under the simultaneity hypothesis (3.1) would not apply before other lenitions (including assimilations) but rather simultaneously. This seems to be borne out in many examples; the following is typical.

- (24) The assimilation of vowels to [ŋ] (9), as in [bæɛŋ] *bang* (which does not apply before [k], e.g. [bæk] *back*), is not bled by nasal elision, e.g. [bæɛŋk] *bank*, not \*[bæɛk].

Perhaps this is the reason for the different behavior of insertions and deletions that Kenstowicz and Kisseberth observed.<sup>20</sup>

**3.2.2 Rules first, processes last.** Phonological rules (2.5), or, for that matter, all rules of language (syntactic, morphological, secret language, etc.) which are not phonetically motivated, apply before phonological processes. This is a traditional hypothesis, except in the standard generative literature. It is abundantly evidenced, and its phonetic teleology is self-evident. The principle resolves a number of otherwise problematic cases. For example, the rules deriving *obscenity* and *dreamt* from *obscene* and *dream*, apparently lenitions, bleed the fortition processes diphthongizing tense vowels, e.g. [i] → [ij]. But they are *not* lenitions, because they lack any contemporary phonetic motivation. The same rules feed the dialectal [ɛ]-raising before nasals which we described as counterfed by all other processes, e.g., *obsc[i]nity*, *dr[i]mt*.

To take a less obvious example, the contraction of *it is* [ɪtɪz] to [ɪts] *it's* (presumably via [ɪtʒ]) bleeds flapping ([ɪtɪz]), rather than applying simultaneously to give \*[ɪtɪz]. If the rules-first principle is correct, it is probably the case that contraction is not a phonological process. We don't really know *why* it isn't, though we could cite a number of excuses. The question, which is at least in part a question of morphosyntactic principles, transcends our subject matter, and our knowledge.

**3.2.3 Remarks.** Most of the examples in the literature conform to these two principles. The exceptions occur mostly in languages for which we have not had access to speakers. In the past, speaker judgments have rarely been elicited on these issues, but without judgments of negative possibilities (like \*[tɪmpθi] for *Timothy*) it is impossible to obtain direct evidence that nonoccurring interactions (like syncope feeding epenthesis) *could* not occur.

The strength of these judgments, and of the independent evidence (2.4–5) on which the rule/process and fortition/lenition distinctions are based, provide strong support for the two precedence principles proposed here. Both principles—that nonphonetic operations yield the last word to phonetically motivated operations and that perceptually motivated operations yield to articulatorily motivated ones—have straightforward phonetic teleologies, and therefore might ultimately provide something more than a description of the facts.

**3.3 On constraints.** In 3.1 counterfeeding was treated as a counter-

iteration constraint. Whenever its expository value, counter-iteration is *not* an independently evidenced constraint. Self-feeding processes never are constrained to apply to just one of a string of susceptible segments, e.g. \*[hænbɪkt] *hand-picked*, \*[meɪnli] *mainly*. Rather, like all processes, their scope is limited by *phonetic*, including prosodic, conditions.<sup>21</sup> For example, sonorant nasalization (1, 7) is halted only by a nonsonorant segment, or by a syllable or accent-group boundary (2.2).

Counterfeeding is a constraint on the derivational status of susceptible representations. Whereas in feeding (unconstrained application) a process applies to *all* susceptible representations, and in suppression it applies to *none*, in counterfeeding the process applies to all susceptible basic representations but to no derivative ones. If we consider what this implies about a speaker's phonetic capacity, it is easy to see why, as we have argued in 3.1, a process which is counterfed by one process should be counterfed by others. Having learned not to flap an intervocalic [t] in [plæɪtɪ] *plant it*, one has acquired the capacity to pronounce it in [feutɪt] *felt it* as well.

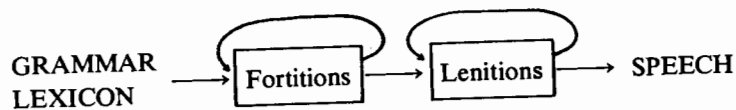
The question which remains conspicuously unanswered, however, and which is surely behind the lingering skepticism about counterfeeding even in the face of numerous examples, is why, if one can pronounce derivatively intervocalic [t] in *plant it* and *felt it*, should one find it difficult to pronounce basically intervocalic [t] in *hit it*? In fact, mastering a derivative representation ordinarily *does* enable one to pronounce the corresponding basic representation. Speakers who pronounce intervocalic [t] in *plant it* and *felt it* ordinarily can also pronounce it in *hit it*, for example to distinguish this from *hid it*. But such pronunciations are occasional at best, and require special effort which the counterfed representations do not.<sup>22</sup>

The reason, we think, is that at the margins of competence, it is easier to achieve a specific objective by aiming at an objective whose difficulty transcends that of the original. This principle is well known to anyone who has done anything difficult, from playing piano scales rapidly to removing a stuck jar-lid. Consider the example of flapping. If the average American speaker *aims* to say an intervocalic flap, e.g., in imitating the British pronunciation of *hear it*, he usually fails. If he aims at an intervocalic /t/, on the other hand, he normally achieves the flap [ɾ], although even here he occasionally deletes it altogether. If he aims at a long intervocalic /t:/, as in Italian or Japanese, he may achieve a simple [t], though, as teachers of these languages can attest, he occasionally only achieves a flap; a total deletion of this target, however, is not so likely.

Translating this last example into occurring English representations, if the target is /VntV/ or /VltV/, an intervocalic [t] is more likely to result than if the target actually is /VtV/.

This 'Excelsior!' principle is clearly what is, in part, behind such recurrent fortitions as (a) [aɔ] → [aɔ̃] (e.g. *cow*, general US) → [æɔ̃] → [eɔ̃] (recent Baltimore); (b) [æɪr] → [aɪr] (e.g. *ire*, general US) → [aɪə] (southern US lowlands) → [ajə] (Faroese); (c) [nd] → [nt] (e.g. Yiddish *hint* 'dogs') → [nts] (Bantu), etc. The likelihood of the lenitions (a) [aɔ] → [ɒ] (e.g. *law*, early English), (b) [æɪr] → [ar] (southern US highlands), or (c) [nd] → [nn] (occasional US *candy*) → [n] (*and*) is lessened by increasing, through fortition, the distance between the actual target and the shortfall lenition product. This 'prophylactic' strategy, which students of sound change from Grammont through Martinet have recognized, is another reason why, as we argued in 3.2.1, even synchronic fortitions precede lenitions. To keep *bands* /bændz/ distinct from *bans* /bænz/, we bleed the deletion of [d] by prior fortition of [z]: [bændz]; to keep [aɔr] *our* distinct from *are* /ar/, we bleed the applicable reductions by prior fortition of [o] and [r]: [aʊr]; and so forth.

It is time to summarize. The model of the natural phonological system presented in 3 can be diagrammed thus:



There are, broadly, three degrees of extrinsic nonphonetic constraint on processes: application to *any* susceptible representation, application to *none* (suppression), and application just to basic representation (counterfeeding).<sup>23</sup> The diagram reflects our conclusion (from examples like (15–17) in 3.2.1) that lenitions cannot be prevented from applying to the output of fortitions.

#### 4. REPRESENTATION.<sup>24</sup>

We now consider, very briefly, the phonological representations which underlie our speech, which we perceive as underlying the speech of others, and which we commit to memory as the phonological forms of the words of our language. We have argued that the phonological processing of what we say is governed by phonetic teleologies. It can be reasoned that the processing of what we hear is a (subconscious) form of teleological anal-

ysis, projecting from what is heard to the phonological intentions of the speaker. This analysis is carried out (with some adaptations to the speaker in question) through the same system of processes that governs our own speaking. This is evident from the fact that, when we listen to our own speech, what we perceive is not what we actually say, but precisely what we intend to say.<sup>25</sup>

Although Sapir (1921, 1933) pointed out that this intended representation is far more readily brought to consciousness than the 'actual rumble' of speech, it is remarkable that half a century later there is little agreement about the character of phonological representations even in the language of the majority of the world's linguists.

One reason is that the search for phonological unity underlying the superficial phonetic variety of speech started, from the beginning, from two distinct points of departure—phonetics and grammar—and arrived at two distinct (although, before the thirties, not clearly distinguished) conclusions—phonemic and morphophonemic representation. The structuralists approached language 'inductively', from the hearer's (or even the learner's) vantage, and neglected (or even rejected) morphophonemics. The generativists have approached language 'deductively', from the speaker's vantage, and have rejected phonemics.

The other reason for the current disagreement is that in both structural and generative phonology, phonological representation has been treated more as a device for simplifying and generalizing phonological descriptions than as an empirical hypothesis. Until the past decade (Stampe 1968a, Kiparsky 1968a, b, etc.—see Zwicky 1972b for references), evidence independent of the facts under description was virtually never cited. Sapir's psychological reality paper (1933) was ignored by structuralists who rejected morphophonemics, and Twaddell's Old High German umlaut paper (1938), for example, was ignored when the generativists rejected phonemics.<sup>26</sup> From the hearer's vantage, [rɛ̃ɪrɪŋ] is unambiguously analyzable as [rɛ̃ɪrɪŋ] *raining*,<sup>27</sup> whereas [rɛ̃ɪrɪŋ] is ambiguous between [rɛ̃ɪrɪŋ] *raiding* and [rɛ̃ɪrɪŋ] *rating*, in the absence of further, nonphonological information. A purely phonological, or *phonemic*, underlying representation of [rɛ̃ɪrɪŋ] and [rɛ̃ɪrɪŋ] would be /rɛ̃ɪrɪŋ/ and /rɛ̃ɪrɪŋ/, respectively, since all English intervocalic flaps derive unambiguously from stops. But whereas the nasal flap can only derive from a voiced nasal stop (since English lacks voiceless nasal stops), the non-nasal flap might derive from either a voiced or voiceless stop, both of which occur in English and both of which become voiced flaps between vowels (process (6)). Phonological representations distinguishing

|rejdɪŋ| and |rejtɪŋ| are called *morphophonemic* because they incorporate information derived from other pronunciations of the respective morphemes, e.g. [rejdz] *raids* versus [rejtɪs] *rates*. Thus *rating* has different representations at three 'levels': phonetic [rejdɪŋ], phonemic /rejdɪŋ/, and morphophonemic |rejtɪŋ|. Structuralist grammars typically treated phonemic and morphophonemic representations separately.

This English example 'translates' a Russian one from which Halle (1959) argued that incorporating the phonemic level merely complicates a phonological description, in that it may require a single process (flapping) to be separated into phonemic ([n] → [ɾ]) and morphophonemic ([t] → [ɾ]) parts. Actually the English example is even more complicating than Halle's Russian one. The immediate output of flapping applied to [t] is not actually [ɾ] but the voiceless [ç], which surfaces occasionally in word-final contexts (e.g. [rejtɪç] *rate*), but which is obligatorily voiced between voiced segments. Flapping [t] → [ç] is therefore a phonemic process, while flap-voicing [ç] → [ɾ] is morphophonemic (it merges [rejtɪçɪŋ] with [rejdɪŋ]). But the phonemic process applies *before* the morphophonemic process,<sup>28</sup> and therefore the phonemic representation of *rating*, /rejdɪŋ/, does not arise at *any* step in its processing.

But Halle attacked a straw man. Phonemic descriptions were never process descriptions, and they made no attempt to describe the relations between levels, even *within* the phonemic component of the description, in terms of processes. Typically the realizations of one phoneme were listed with their respective contexts (e.g., /n/ = [ɾ] between vowels, [-] before voiceless homorganic stop, etc.) without systematic cross-reference to the realizations of other phonemes (e.g. /d/ = [ɾ] between vowels, [ɔ] before /θ/, etc.), regardless of parallels. This is not to say that phonemicists were unaware of processes. Allophones arising from the same processes were usually described in the same order and the same wording under their respective phonemes. But there is nothing in phonemic theory or practice to suggest that a single process might not govern alternations of allophones in one case and phonemes in another. For example, the whole point of Twaddell's highly regarded phonemic analysis of Old High German umlaut (1938), was that this process was phonemic in some applications but morphophonemic in others. Halle's argument is totally irrelevant to the status of phonemic representation.

In fact, it is only in the generative theory of 'systematic phonemics' (Halle op. cit., 1962, Chomsky 1964, Chomsky and Halle 1968) that phonological representations are supposed to correspond to a specific point in a list of ordered processes. According to this theory, phonological

representations arise after the application of 'morpheme structure' (or 'phonological redundancy') rules, which govern basic representations, and before the application of proper 'phonological rules', which govern alternations. We have already mentioned examples which falsify this. The constraint against intrasyllabic clusters of stop plus nasal (\*|bnrk|) is due to the obligatory syllabification of the nasal, by the same process that accounts for optional syllabic alternants of liquids or glides, e.g. [blʌri] ~ [b|ʌri] *bloody!*. The constraint against [h] before consonants (cf. OE *hnutu*) is due to the obligatory deletion of the aspirate by the same process that accounts for optional alternants like [hjuj] ~ [juj] *huge*.<sup>29</sup> Further, the constraint against intrasyllabic clusters of voiced and voiceless obstruents (\*|sgpɔd|) is due to the same processes that assimilate voicelessness in alternations like [Its gɔn] ~ [skɔn] (*It's gone*, [ɾɪbd] *ribbed* but [ɾɪpt] *ripped*, etc. What these examples show is that some processes that govern phonological representation also govern phonetic representation.

There is other evidence against the systematic phonemic theory. Note that according to the conception of levels as corresponding to a point in an ordered list of rules, all morphemes would have phonological representations at the same point in the list of processes. For example, in German, since [ve:k : ve:gə] *Weg* : *Wege* 'road : roads' and [vek : vekə] *Weck* : *Wecke* '(breakfast) roll : rolls' require their phonological representations to be 'prior' to the process devoicing final obstruents, [vek] *weg* 'away' also requires a representation prior to devoicing. Since the latter is uninflected, there is no way to determine whether its final obstruent is phonologically voiced, with devoicing by the devoicing process, as in *Weg*, or whether it is phonologically voiceless, as in *Weck*. Therefore the obstruent of *weg* is supposed to be represented as an 'archisegment' phonologically unspecified for voice. W. S.-Y. Wang and Stampe (1967, oral interventions on Kiparsky 1968a) pointed out that the development of Eastern Yiddish (Sapir 1915), wherein the devoicing process ceased to apply, argues against this. Medieval German *wec* (*Weg*) and *onwec* (*weg*) became Yiddish *veg* and *avek* and, in general, forms with invariably final obstruents were treated exactly like forms where these alternate with nonfinal voiceless obstruents. The archisegment hypothesis suggests, incorrectly, that they should have become randomly voiced or voiceless.

The idea of incompletely specified phonological segments is a persistent recurrence in descriptive phonology, from Twaddell's 'macro-phoneme' (1935), Trubetzkoy's 'archiphoneme' (1969), through Jakobson, Fant, and Halle's blanks in feature matrices (1951), down to



generative archisegments (Chomsky and Halle op. cit., Hooper 1976). It began with the structuralist definition of phonemes, according to Saussure (1949), as opposite elements, the sum of their *distinctive* features (Jakobson 1932, Bloomfield 1933), and the observation that in positions of 'neutralization' by this definition there are segments of dilemma, neither distinctively *x* or non-*x*. For example, beside the distinctively voiceless and distinctively voiced stops of *pin* and *bin*, there is the stop of *spin*, which is not *distinctively* voiceless because \**sbin* is not admissible. In generative phonology, this problem was simply transferred from phonemic to morphophonemic representations. Only Twaddell seems to have grasped that neutralization is a refutation of the idea that phonemes are opposite elements.

The single argument that is offered for archisegments—uncertainty—has about as much force as a blindfolded man arguing that it is neither night nor day (or that it is both) because he can't see which it is.<sup>30</sup> In fact there are many ways to ascertain how speakers evaluate such segments. It might be noted that no alphabet provides special symbols for archiphonemes distinct from phonemes. Or that archiphonemes rhyme better with one phoneme than the other: for example, *spin* alliterates perfectly with *s'pose* but not with *s'bbatical* even if they are pronounced alike with [sp]. Or that when the cluster is split up by epenthesis or prothesis in children (e.g. [sikul] *school*) or in historical change (e.g. Spanish *escuela*), the stop, removed from the devoicing influence of the [s], shows up as voiceless.<sup>31</sup> Of course, this is precisely what is predicted from the fact that English has a process obligatorily devoicing stops after tautosyllabic voiceless segments, as in the example [skɒn] (*It's gone*) cited above.

From this example and the example of Yiddish *avek* one might conclude that all processes governing phonetic representation also, in the absence of motivating alternations, govern phonological representation. This would amount to claiming that the basic level of phonological representation is the phonetic level.<sup>32</sup> But there is much evidence against this. For example, the arguments against the archisegmental evaluation of stops after tautosyllabic /s/ show that speakers do not perceive these as a third phonological value distinct from both initial voiceless and voiced stops. But they also show that invariant phonetic values are not necessarily phonological, because stops after /s/ do have, in English, a third phonetic value distinct from those of initial stops. Speakers are, however, totally unconscious of the difference between e.g. initial [k<sup>h</sup>] and non-initial [k], even in alternations like *crunch* : *scrunch*, *it's cold* : *'s*

*cold*, etc. Sapir (1933) provided a similar argument when he pointed out that his Nootka guide wrote *hi*, *hu* for the invariant syllables [hɛ], [hɔ], disregarding the lowering of the vowels by [h]. Therefore, if Sapir's characterization of phonological representation as that which is most readily 'brought to consciousness' is accepted, as we think it must be if the notion is to have any psychological significance, we must conclude that many phonetic features of speech, even though invariant, find no place in our phonological consciousness or memory.

But *which* invariant features are nonphonological? We have already seen that the conventional answer to this question in both structural and generative phonology, that it is the redundant features which are absent from phonological representations, is incorrect: stops after tautosyllabic /s/ are perceived as voiceless despite the fact that this voicelessness is predictable and nondistinctive. Conversely, the vowel of e.g. [kæ̃t] *can't* is perceived as nonnasal even though it is distinctively nasal (contrast [kæt] *cat*). The distinctiveness principle fails left and right. What alternatives are there?

With Sapir, we could understand phonological representation to be the phonological intention (and perceived phonological intention) of speech. We have characterized the natural phonological system as the system of limitations which stand between the intention and the actualization of speech—i.e., between phonological and phonetic representation. The principle of phonological perception must be *naturalness*: if a given utterance is naturally pronounceable as the result of a certain intention, then that intention is a natural perception of the utterance (i.e., a possible phonological representation).

The utterance [spɛ̃t] will illustrate this naturalness principle. We can perceive this as [spɛ̃t] because if we pronounce [spɛ̃t], what we actually say is precisely [spɛ̃t]—nothing in our acquisition of English has taught us *not* to nasalize vowels before nasals (1), or *not* to delete nasals before homorganic voiceless consonants (5). We cannot simply perceive the utterance as [spɛ̃t] because, if we tried to pronounce [spɛ̃t], what we would actually say is [spɛ̃t]—nothing in our acquisition of English has taught us to pronounce vowels as nasalized on purpose. Our English phonological capacity is dominated by a fortition process which denasalizes all nonstopped segments, including, as in the present instance, vowels. This tendency to denasalize is well known to those who teach French or Hindi to English students, and its natural character is attested by its occurrence in children (e.g., Joan's [ɑts] *ants* [Velten 1943]) and in historical change (e.g., the loss, in Icelandic, of the nasalization of vowels

recorded by the First Grammarian in words like *í* 'in'). We do not denasalize the [Ē] of *spent*, of course, because fortitions do not follow lenitions (3.2.1).<sup>33</sup>

But why don't we perceive [spēt] as |sbent|? After all, there is a process which devoices stops after tautosyllabic voiceless segments, and in fact if we try to pronounce |sbent| what we naturally say is [spēt]. But while we cannot intentionally pronounce [Ē], English has taught us to pronounce [p], and therefore we have no reason not to take the [p] of [spēt] at face value.

Or consider the example [bæ̃r̃f̃]. Its face value representation |bæ̃r̃f̃| would be pronounced [bæd̃f̃], since nonstops (including flaps) are denasalized, as noted above, while flaps are (simultaneously) stopped. (This fortition is heard in American attempts at the initial [r] of Japanese; it is also heard in children [Edwards 1973: appendix] and in historical change, in fact in some Japanese dialects.) To obtain [bæ̃r̃f̃], therefore, we must aim at |bæñf̃|, which regressive and progressive nasalization (1,7) and flapping (6) convert to [bæ̃r̃f̃].

These examples minimally illustrate how, in the analysis of utterances, the naturalness principle establishes a basic level of phonological perception, distinguishing features which are phonological from those which are merely phonetic. This level corresponds closely to the phonemic level the structuralists sought to capture.<sup>34</sup> It is an instructive exercise to seek out, in the intrinsic restrictions imposed by the natural phonological system, the natural analogues of structuralist analytic criteria, concepts of markedness, implicational laws, and so forth. It is also instructive to compare natural phonemic analyses, which are typically unique, with the alternative analyses the structuralists debated for English diphthongs, Spanish glides, or more recently, Kabardian vowels. But this must await other times and perhaps other authors.

Here we must be content to point out that, as our examples show, only sounds which pass the muster of the obligatory fortition processes of a language are phonemes. The remainder, optional or lenition-created variants ('allophones') of the phonemes, play a role only in the subconscious aspects of perception, and therefore find no direct representation in our morphophonological memories, our formulations of phonological or grammatical rules, our spelling systems, our verbal play, or even our lapses of speech or hearing.

Some of these results are implicit in evidence discussed, to other ends, above; and some are illustrated in Stampe (1973a). Here we will cite just the example of rhyme (Stampe 1968a), which (like alliteration,

Icelandic *hendingar*, Welsh *cynghanedd*, etc.) requires phonemic identity. If pronounced with phonemic identity, *fitter* : *bidder* : *hit 'er*, *scan it* : *plant it*, *fans* : *hands*, *expense* : *rents*, *hen* : *tend*, *stole* : *old*, *mix* : *sticks* : *sixth(s)*, *mess* : *pest* : *tests* : *desks* are rhymes, despite their morphophonemic differences. (Of course, rhymes due to casual neutralizations, as in *hen* : *tend*, naturally sound correspondingly casual, or even silly, beside those due to obligatory neutralizations, e.g. *hens* : *tends*.) And unless morphophonemically identical words are pronounced with phonemic identity, they do not rhyme: *rolled* : *ol(d)*, *twined* : *rin(d)*, *plot them* : *got (th)em*, etc. Phonetic identity, moreover, is entirely irrelevant: in *There was a gnat/ upon a cat/ upon a pad/ upon a mat*, *gnat* and *pad* do not rhyme even if pronounced as [næɹ] and [pæɹ],<sup>35</sup> whereas *gnat* and *cat* rhyme perfectly even if pronounced distinctly as [næɹ] and [kæt] or [kæʔ]. Only the phonemic identities matter.<sup>36</sup>

Morphophonemic representations certainly exist, of course. [spēt] can be perceived, without violations of the naturalness principle, as |sbent| (*It*)'s *bent* as well as |spent|, and [bæ̃r̃f̃] can be perceived as |bæñ h̃f̃| *ban her* or |bæ̃nt̃f̃| *banter*. Such perceptions are motivated when a form's various pronunciations are not collectively derivable, through the natural processes of the language, from their phonemic representations. The utterances [téligræf] *telegraph* and [tílégrifi] *telegraphy*, phonemically /telagræf/ and /tálégrafi/, require the morphophonemic |teleg̃ræf| to derive both from a single representation through the natural process of vowel reduction. The 'depth' of such representations is an idiosyncratic matter, as we have argued earlier, varying from form to form.

Phonological representation is best understood not as a level but as a *kind* of representation, namely the representation of forms in permanent memory. With this conception it is easier to understand why phonological representations do not incorporate sounds beneath the level of phonological perceptibility, the phonemic level as defined by the naturalness principle.

The importance of the phonemic level is reflected in a variety of subtle ways: in the fact that we say that *banter* or *ban her* are pronounced like *banner*, not vice versa; that eye dialect spells *for*—[fɔɹ] ~ [fɹ], the latter phonemically /fɔɹ/—like *fur* [fɔɹ], not vice versa; and of course, by the gradual historical replacement of morphophonemic representations (|test| for [tes] *test*, due to plural [tes] *tests*) by phonemic ones (dialectal [tes], compare the plural [tesiz]).

These conclusions flatly contradict Chomsky's claim (1964) that phonemic representation is without linguistic significance, and cast doubt

on his claims for a level of systematic phonemics. There is one further claim to examine: that phonetic and phonological representations are mediated not only by natural processes, but also by rules in the strict sense of 2.5. We do not hope to *settle* this issue, because generative phonologists have provided no explicit empirical characterization of systematic phonemics. However, since Halle (1959) and Chomsky (1964) claim that it is close to Sapir's phonological representation, we believe it is reasonable to expect that systematic phonemics too should be more readily available to consciousness than other representations. This expectation seems particularly reasonable since systematic phonemic representations like  $|\text{de} = \text{k}\bar{\text{i}}\text{d} + \text{iVn}|$  *decision* and  $|\text{ærtifik} + \text{i} + \text{æ} + \text{i} + \text{ty}|$  *artificiality*, and the rules which relate them to phonetic representation, obviously presuppose operations which are more *cognitive* than the phonemic and morphophonemic representations we have been discussing. However, we doubt whether Sapir's empirical criterion would be accepted, because when independent evidence like Pig Latin (Halle 1962) or metrical scansion (Kiparsky 1972) has been examined, it has been concluded that these interact with phonological derivations at 'natural breaks' somewhere midway in the list of ordered rules. No general characterization of such 'breaks' has been offered, and no explanation of why such 'breaks' should occur at midpoints rather than at the systematic phonemic level.

The fact is that no independent evidence of *any* kind (a list of kinds and references is provided by Zwicky 1972b) requires a systematic phonemic explanation.<sup>37</sup> Secret language rules, for example, provide one kind of independent evidence for phonological representations—as when infixing secret languages like Alfalfa or Ob allow recovery of neutralized natural phonological representations by blocking the neutralizing process:  $[\text{plælfintælfɪ}]$  *planter* versus  $[\text{plælfɪnælfɪ}]$  *planner* (infixes italicized). But systematic phonemic representations, e.g., the  $|\text{d}|$  posited for  $|\text{z}|$  in *decision*, or the  $|\text{k}|$  posited for  $|\text{s}|$  in *electrician*, fail to turn up:  $[\text{dabɪsabiʒabɪn}]$ ,  $[\text{abilabɪktrabɪʒabɪn}]$ . We have seen that rhymes, although basically phonemic, are much better with (morpho-)phonological identity. We would expect this also to be true of systematic phonemic identity. But it is not: *decision* (*decide*) rhymes perfectly with *revision* (*revise*) and *precision* (*precise*); and so it goes for *extension* (*extend*), *retention* (*retain*), *convention* (*convene*), and *tension* (*tense*); for *resign* (*resignation*) and *incline* (*inclination*); for *meant* (*mean*), *bent* (*bend*), and *tent*. Finally, the enormous difficulty phonologists have discovering systematic phonemic representations and the kind of phono-

logical rules they require, *even in their native language*, hardly squares with Sapir's criterion of accessibility. In short, we see no reason whatever to believe that phonological representations are motivated by phonetically unmotivated rules.

## 5. FINAL REMARKS

In the previous section of our paper, we invoked an empirical criterion of Sapir's on the issue of representation. It is remarkable that we should have to reach back into traditional phonology for an empirical criterion after half a century of empiricist theories. But the fact is that although structural and generative phonology are empiricist they are not empirical. Chomsky based his critique of structuralist phonology on the fact that structuralist theory did not define the particular sorts of representation which were generally agreed to be phonemic. If there had not been this fortunate agreement that, e.g., *can't* is phonemically  $/\text{kænt}/$  rather than  $/\text{kæt}/$ , Chomsky's critique would have been impossible. The structuralists had provided no independent empirical characterization of phonemic representation: they did not say what it is supposed to explain. They did, however, say what it is supposed to be explained by, namely the distinctiveness principle. In generative phonology we have no characterization of either what is supposed to follow from the theory, or what it is supposed to follow from. It is neither falsifiable, on the one hand, nor explanatory, on the other.

This is apparent from the fact that when any particular aspect of either structural or generative phonology has been falsified by data, either the data have been declared irrelevant, or the hypothesis has simply been revised, and the respective theories have gone on their way unruffled. When the distinctiveness principle in structural phonemics confronted the problem of neutralization, some structuralists declared the problem of identifying the phoneme in the position of neutralization irrelevant, and others simply changed to alternative principles. Generative phonologists, confronted by difficulties with the feature-counting principle originally proposed as an 'explanatory theory' of phonological representation, either abandoned the problems it had been proposed to explain (Chomsky and Halle 1968: chapter 8), or turned to alternative principles like markedness (op. cit.: chapter 9) or other equally unrelated criteria (Zwicky 1972b provides twenty-six criteria from the literature). In other sciences, the abandonment of such basic goals or principles would be revolutionary.

In structural and generative linguistics, they have occurred with less notice than is accorded a change of notation. For all their rigor and explicitness, neither structural nor generative phonology has *essential* empirical content. They are, to put it simply, not theories.

Natural phonology, although it lacks any a priori methodology or formalization, is both testable and explanatory. By its nature, it is ultimately accountable for, as we put it earlier, everything language owes to the fact that it is spoken. And by its nature, it must follow from the character of the human capacity for speech.

It has been objected that too little is known about phonological universals and about phonetic capacity (especially in its neurological aspect) to falsify the theory. Even if we accepted this assessment of the phonological and phonetic literature, it would not follow that the theory is unfalsifiable *in principle*. (In fact, however, we think that the literature has already proven adequate to support systematic investigations of many aspects of the relations between phonology and phonetics.)

Others have objected that the theory is too obviously true to be falsified. We can only conclude that this objection is based on an unawareness of the intricate, complex, paradoxical, and nonpatent nature of the facts to be accounted for. In any event, if it is obviously true, it is certainly not obvious why the theory has lain dormant for a half-century.

In the meantime, the goals of explanation which were set by the pioneers of phonology and phonetics referred to at the beginning of our paper have largely been forgotten, along with the considerable progress they made in achieving these goals. In their place we have, as the late Paul Goodman wrote of modern linguistics in general, 'an enormous amount of machinery, but few edible potatoes.' We hope we have been able to show in this paper that a return to the traditional goals may increase our yield.

## NOTES

This paper is dedicated to the memory of Harry V. Velten of Indiana University, whose studies have lighted our way.

1. There are by now other more or less independent varieties of the theory current, and most recent revisions of generative phonology have converged with natural phonology. There are also divergent views within the various schools, even between the co-authors of this paper. The common

ground is the basic thesis that phonological systems are phonetically motivated.

2. That is, obstruents ([z]) become voiceless ([s]) in final position. The reverse is not true, e.g. *weiss* : *wissen* '(I) know' : '(they) know' retain [s] throughout. Although Kruszewski and Baudouin recognized that alternations are unidirectional, and that one alternant is basic and the other derivative, they avoided interpreting them as processes because of the diachronic overtones this notion had in their time.

3. The hypothetical example is superfluous: with a whole world of languages at hand, one issue after another in the book is 'left open', because of 'a scarcity of data for choosing between the many alternatives that readily come to mind' (379 and *passim*). This suggests that the issues are really pseudo-issues. A theory for which one can find no evidence is, in effect, a theory with nothing to explain.

4. For this informal explanation to hold water, of course, it needs caulking with perceptual and articulatory substance. The objective measure of perceptual and articulatory difficulty is necessarily one of the main goals of natural phonology. However, it is a goal presenting enormous obstacles, not the least of which is that even articulatory difficulty seems usually to be as much mental as physical. Therefore, at the risk of some misunderstanding (e.g., Ohala 1974), we employ the notions of perceptual and articulatory difficulty as interim hypothetical constructs, deduced from the rich evidence furnished by the nature and frequency of substitutions in phonological variation, acquisition, and change. This is not circular, because of the coincidence of conclusions drawn from quite different kinds of evidence, e.g., the greater difficulty of perceptibly rounding low as against high vowels is independently attested by their consistently different behavior in a wide variety of substitutions (e.g., Donegan 1973a, 1976, 1978). In fact, it seems to us that natural phonology furnishes systematic data on the nature of features, sounds, and sound-structures which are otherwise unavailable, though indispensable, to linguistic phonetics.

5. Except for some Munda languages, with both voiceless and voiced obstruents, which permit only the latter finally in morphemes; they are pronounced as voiced when syllable-initial (e.g. before a vocalic suffix) but are checked and usually devoiced when syllable-final (Stampe 1965: 333 f. on Sora).

6. Chomsky and Halle (1968) present two revisions of the theory, the first (ch. 8) viciously circular, the second (ch. 9) failing to distinguish admissible from inadmissible *segments* (Stampe 1973c). No explanatory theories of admissibility were proposed by the structuralists, perhaps because (as Chomsky 1964 proposed) they never aspired to this level of "adequacy."

7. Donegan and Stampe (forthcoming). The main process governing this restriction seems to be one which syllabifies the second segment, as in [bɪɪk] or [bɪɪk]; the expressive [bɪu] *bl-lue!*, [kuɪt] *qu-it!* seem to derive from optional applications of the same process—(3f) of section 2.4.

8. There is of course rich evidence that the child's mental target resembles adult speech rather than her own (e.g., Stampe 1969, 1972b, Smith 1973,

etc.). In Joan's case this came in the spontaneous and across-the-board appearance of vowel length and, shortly afterwards, voicing in precisely the words which end in voiced obstruents in adult speech.

9. In this section we sketch a theory differing from our oral paper in two respects, (1) the assumption of simultaneous application (Donegan 1974) rather than freely sequential application (Stampe 1973a), (2) the extension of the 'fortition first, lenition last' hypothesis from pairs of processes with opposite effects (Stampe 1973a) to all processes. These minor revisions have far-reaching consequences, not all of which we can evaluate here.

10. E.g., Koutsoudas et al. (1971), Kiparsky (1971), Anderson (1974), Hooper (1976), etc.

11. Parenthesized conditions are variable. On regressive nasalization cf. 2.2.

12. We have skipped over some steps in the 'spread' of nasalization and have presented nasalization and nasal-elision as simultaneous.

13. In the case of Vennemann and Hooper, the grounds seem to be 'concreteness' for its own sake. They identify processes with empirical statements, expecting them to express true generalizations about phonetic representation. This criterion can be met only at the expense of reducing phonology—which after all has a perceptual side—to articulatory phonetics. To hug the phonetic ground closely is not necessarily to embrace the truth.

14. However, in a non-ordering framework, Lee (1976) has anticipated the observation which follows.

15. The same is true of the arguments of Kiparsky (1972) and Anderson (1969, 1974) for an 'elsewhere' or 'disjunctive ordering' principle which deals mainly with cases where the effects of a specific rule would not be manifest unless it is appropriately ordered with respect to a contradictory and more general rule. Stampe (1973a: ch. 2) has given counter-analyses to Anderson's examples, which, despite Anderson (p. 103, note 8) are abundantly documented—in fact in the handbooks Anderson cites (note 5 of p. 102), to which may now be added Jordan (1974: §25)—and need not be rehearsed here. Kiparsky gives the self-evident logical argument for the principle, citing ancient Indian authority, including Pāṇini's similar convention, but it is obvious that such an argument is completely inapplicable if we are dealing with phonetically motivated processes rather than rules formulated by the learner.

As to Pāṇini, it is surely anachronistic to interpret the descriptive conventions of the *Aṣṭādhyāyī* (as also in Kiparsky's paper in this volume) as if its author had been a conventionalist, i.e., as if the conventions were intended as hypotheses in a theory of language; even Patañjali's exhaustive commentary gives no hint of theoretic rather than descriptive intent. As Stampe remarked at the conference, rules in feeding, bleeding, disjunctive, and cyclic application can be found in any complicated set of instructions, like *The Joy of Cooking*, but one does not interpret these as part of a universal human *faculté de cuisine*.

16. Jensen and Stong-Jensen (1976) point out that many prosodic processes, e.g. alternating stress or length, bleed themselves: application on the

*n*th syllable will bleed application on the *n*th  $\pm 1$ . These examples disappear in a prosodic theory which treats the alternating prominences as part of a prosodic pattern, and maps the segmental material onto this pattern (2.4).

17. In a narrow description of non-southern English, it might be overlooked that these dissimilations of [æɛ] to [œ] exist, since there is nothing for them to apply to. On our view, this only suggests that there is no reason for them to have been suppressed. And in fact they regularly show themselves in northern imitations of the southern pronunciations [bæɛd], [græɛs] as [bœd], [grœs]. One 'dictionary' of southern speech for northerners glosses *died* as 'father'.

18. Lest it be doubted that the lenitions in (10) or (22) are living processes, speakers with these processes often apply them to the outputs of the lenition desyllabifying pretonic sonorants (31), e.g. [θrou] *Thoreau* → [θrou] → [θrou] or [trɪfɪk] *terrific* → [trɪfɪk] → [čɪɪfɪk], though many also use the counterfed pronunciations [θrou] and [trɪfɪk] as variants.

19. There can be little doubt that the much-debated alternants of these affixes have synchronic phonetic motivation in English, because they are adjusted to fit tongue-slips (Fromkin 1971). The distinction in the spelling of the [z]-suffix between *cats*, *dogs* versus *matches* reflects, we suspect, the phonemic (not morphophonemic) status of the inserted vowel, according to the perceptual hypothesis sketched in 4 (cf. also Read 1975).

20. Strictly speaking, there seem to be no natural insertions or deletions; the former involve 'splitting' segments by dissimilation or assimilation, e.g. [æ] → [æɛ], and the latter are simply complete assimilations, e.g. [[ænt] → [æ̃æ̃t] = [æt] (Stampe 1972a, Donegan forthcoming). Note that if [æ] → [æɛ] were really an insertion, its simultaneous application with regressive nasalization would make [bæɪ] into \*[bæ̃ɛɪ], with non-nasal [ɛ]!

21. Kisseberth (1973) cites Lardil examples of Hale (1973) to argue that two processes, vowel apocope and grave consonant apocope, may evidence a counter-iterative constraint on their application, to prevent [ɲawuɲawu] *termite* (cf. the inflected form [ɲawuɲawu-n]) → [ɲawuɲaw] → [ɲawuɲa] (the correct pronunciation) → \*[ɲawuɲ] → \*[ɲawu]. We have no data on this language, but we suspect, both from the usual pattern of Australian languages and from reduplicative structure of this word, that there is some accent (whether primary or secondary) on the basically penultimate vowel, and that apocope, as is normally the case, does not apply to accented vowels.

22. DS, who pronounces *bad guy* [bæggœ] and *let me* [lemi], finds it quite difficult to pronounce *bag* as [bæɛ] rather than [bægg], or *lemming* as [lemɪŋ] rather than [lɪmɪŋ].

23. We do not have space here, unfortunately, to discuss nonphonetic constraints involving grammatical, semantic, or lexical categories, and frequency, etc. (Stampe, forthcoming).

Our model bears a strong resemblance to traditional practice, e.g., in Sapir and his contemporaries (Kenstowicz 1976), where one finds globally expressed interactional constraints like "inorganic [i.e. derivative] increments and losses have no effect" [on the application of the constrained process]. For



a modern discussion from this point of view see also Kisseberth (1973). We are much indebted to Greg Lee and Don Churma for discussion of various topics in section 3.

24. Based chiefly on Stampe (1968a), unpublished.

25. 'Tongue-slips' are the exceptions that prove the rule: they are perceived as slips because they constitute jumbings of the intention of speech. They arise in the input to the phonological processes, and since the processes operate perfectly, as usual, both in synthesis and (allowing for noise and ambiguity) in analysis, we correctly perceive our resultant utterance (spoken or not) as not corresponding to our original intention.

26. It should be pointed out that Twaddell's paper presented phonemics as an explanation of the spelling of OHG umlaut, rather than presenting the spellings as evidence for phonemics. This is a good example of the way phonemics was taken for granted. Similarly, more has been written on how generative representations of English explain English orthography than on how the orthography supports the representations.

27. For purposes of exposition, we ignore for now the analysis [rɛɪntɪŋ] \**raining*, which is also possible.

28. Against the morphophonemic precedence principle of Dinnsen and Koutsoudas (1975). There are many similar examples: Kabardian [əw] varies allophonically with [o:] but [q'] merges with the distinct phoneme [q'w] before round vowels, including [o:] (Kuipers 1960:24 n.10); Yana women devoiced final vowels and merged voiced consonant phonemes with voiceless ones before the devoiced vowels (Sapir 1929:207); English /l/ is labiovelarized in syllable codas, and many speakers delateralize the resultant segment in certain contexts, merging it with [ɹ] as in [haɹ] *how*, *howl*; etc.

29. Actually, this example is strictly speaking not a phonological redundancy rule, because such rules are supposed to supply redundant feature values, not delete, insert, or change segments. However, any such formulation of the constraint against [hn] would imply, incorrectly, that if confronted with a word like *hnutu*, an English speaker who could not pronounce it would change it to something other than [nutu] or [hɲutu]. Halle (oral comment, 1971, on Stampe 1973c) argues that these are considerations of loan phonology and are not relevant to the description of English. But the observable constraints and alternations of English, or any other language, are a subset of the regular substitutions its speakers would impose on unpronounceable words from other languages (cf. Ohso 1972, Lovins 1973).

30. Moreover, the idea is never applied even-handedly: if *sixth* and *sixths* were obligatorily pronounced to rhyme with *six*, no one would represent all [ks] sequences as an 'archisequence' |ks(θ(s))|. The duck/rabbit perceptual phenomenon, incidentally, argues against the archisegment idea; in the ambiguous drawing we see a duck *or* we see a rabbit—not both at once.

31. Stampe (1973a) cites numerous further arguments (cf. Velten 1943). Hooper (1976) cites interesting counterarguments of Blair Rudes' showing that Gaelic takes stops after [s] to be voiced. We do not know why this

should be. But our point is that no language takes such segments to be indeterminate. Gaelic is not exceptional.

It has been argued that *spin* has [b] from the fact that when the [s] of *spin* is removed by electronic mutilation the residue is heard as *bin* rather than *pin*. The reasoning here is on a par with claiming that lizards are snakes because if you cut off their legs people will think that they're snakes.

32. This is the simplest hypothesis, given the evidence cited. Stampe presented it, with the counterevidence that follows, in several papers (e.g. 1968a, 1973a), although no one had actually espoused the hypothesis. But now Vennemann (1974) appears to have done so.

33. Joan Velten's denasalization in *ants*, which is regular, might have several explanations: (1) she failed to perceive the vowel nasalization at all (|æts|); (2) she perceived it as phonemic (|æ̃ts|) and applied denasalization; (3) she perceived it as allophonic (|ænts|), but applied denasalization after nasalization and nasal-elision, against 3.2.1. We are not very happy with any of these alternatives. Worse, many children write *ants* as *ATS* (Read, 1975, with important discussion).

34. This claim was anticipated in a remarkable 1954 article by Bazell, in which he argued that phonological identifications are governed not by the principle of (non)distinctiveness but by a principle of *motivation* (essentially our principle of naturalness). We do not identify [h] and [ɪ] in English, Bazell says, even though they are not distinctive, because pronouncing /ɪ/ initially as [h], or /h/ finally as [ɪ], is not motivated.

35. A fluent reading eliminates the quantity difference.

36. The omnipotence of the word asserts itself in the perfect rhymes of *fitter : bidder : hit 'er* (with clitic *her*) versus *a gnat upon : a pad upon*. The former are invariably pronounced alike and thus are phonemically identical; the latter, even in this context, are only facultatively alike (contrast [ənæt'] : [əpæ:d]) and thus are phonemically distinct.

37. Aspects of English orthography which purportedly require a systematic phonemic explanation are better explained historically.