

# ISSUES IN PHONOLOGICAL THEORY

*Proceedings of the Urbana Conference on Phonology*

*edited by*

MICHAEL J. KENSTOWICZ

*and*

CHARLES W. KISSEBERTH

*University of Illinois*

1973

MOUTON

THE HAGUE · PARIS

## ON CHAPTER NINE\*

DAVID STAMPE

In his monograph *Kindersprache*, Jakobson showed that the values of features are asymmetrical, in that in a given segment one value is expected, or 'unmarked', and the other, if it occurs at all, is less expected or 'marked'. From this discovery Jakobson formulated his universal laws of implication as the foundation of phonology. Stated generally, these laws have the form 'If a language admits the marked value of a feature, it admits also the unmarked value'. In vowels, for example, nasality is a marked value, and languages which admit nasalized vowels universally admit non-nasalized vowels.

Jakobson's interpretation of markedness is the basis for a major revision of generative phonology proposed in Chapter Nine of Chomsky and Halle's *Sound Pattern of English*. It is proposed that universal grammar includes a number of 'markedness conventions' such as

$$\left[ \begin{array}{c} \text{V} \\ \text{uNasal} \end{array} \right] \rightarrow [-\text{Nasal}]$$

which is interpreted: Vowels unmarked for nasality are nonnasal (and hence vowels marked for nasality are nasal). Such conventions are intended to explicate the 'naturalness' of phonological systems, under the assumption that (*ceteris paribus*) the most highly valued system is that which requires fewest marked features. Thus a system with only oral vowels is more natural than one with nasal vowels.

For reasons not clear to me, Chapter Nine does not pursue the implicational consequences of the proposed theory. Jakobson demonstrated not merely that marked values are more complex than unmarked, but also that they IMPLY the (admissibility of) the unmarked. I will return to this below. But although Chapter Nine discusses few empirical consequences of markedness theory it seems that the theory confronts many of the same issues as the theory of natural phonology which I have been working at for several years.

In my opus solum (Stampe 1969) I argued that there are in fact neither implicational laws nor markedness conventions but rather an innate system of phonological

\* This paper has benefitted from discussions with Jim Heringer, Greg Lee, and Bill Darden.



processes which resemble the implicational laws and markedness conventions in content but have the same ontological status as the natural processes (so-called "rules") of the phonological system of any individual language. I proposed that the natural phonological system of a particular language is what remains of the innate system of processes after the learner of the language has suppressed, limited, or ordered these processes in such a way as to render accessible the mature pronunciation of the languages. In unpublished work I have attempted to show that, without auxiliary hypotheses, this view of the phonological system likewise explains the nature and existence of representations deeper than the phonetic level.

The chief difference between such innate processes and implicational laws or markedness conventions lies in the fact that the latter are considered not to be of the same ontology as the processes of a particular phonological system. Rather they are universal laws or conventions which govern and interpret the processes of particular systems. Thus, whereas I view the mastery of nasality in vowels as the suppression (or limitation or ordering) of the innate process denasalizing vowels, in markedness theory it would be viewed as the acquisition of the marked value for nasality in vowels, i.e. as the acquisition of an exception to the law or convention that vowels are expected to be nonnasal. In those theories, the language learner can break a law or flaunt a convention, but he is not permitted to repeal either.

I would like to explore some consequences of the assumption that naturalness is governed by metagrammatical conventions for the claims made in Chapter Nine, since, with a few exceptions, what are presented there as axioms of markedness theory seem to be in fact corollaries of the initial assumption. These particular corollaries embody what I think are totally unsupported claims about the nature of phonology, but the issues are so large that rather than citing a handful of data as counterevidence, I will for the most part depend on the judgment of my audience to evaluate the claims in question.

One issue which any phonological theory has to confront is the distinction between processes which merely govern alternations and those which, whether they happen to be manifested in alternations or not, reflect genuine limitations on what we can pronounce. One of the basic differences between these is that processes reflecting our limited speech capacity – for example, the palatalization by *y* of apical obstruents in English – do not suffer native exceptions, whereas processes which merely govern alternations – like SPE's rule of velar softening – do. The palatalization process is subject to many conditions, but it is equally applicable to every word meeting the conditions. In fact, as I believe is true of all context-sensitive processes reflecting our speech capacity, the application of palatalization is considerably but equitably extended in relaxed or allegro speech. Velar softening, which governs such alternations as the *k/s* of *electric/electricity*, is on the other hand subject to many exceptions (e.g. *persnickity*). It does not represent a limitation of our speech capacity, and therefore such a word as *elektrikity* poses no difficulties for us, and, no matter how we mumble, *persnickity* never becomes *persnisity*. Palatalization may optionally apply to Fig-



Latinized words – e.g. [esyē] ‘yes’ > [ešyē] – but velar softening may not – [ōkyē] ‘yoke’ > \*[ōsyē]. Palatalization applies to nativized loan words, e.g. *Katya*, but velar softening does not. These and many other tests reveal this basic distinction between processes, and demand from our theories some principled account of the distinction. In natural phonology, processes like palatalization are residues of the innate system of processes which governed our infant speech, whereas if velar softening was ever one of these, we were forced to suppress it in order to pronounce *lickity split* and *kittycat*, and thereafter we learned to say *electricity* with [s] not because it was easier to say but because the language as we confronted it required this pronunciation in this context. That is, if velar softening is a process at all, it is one we learned.

Markedness theory similarly makes a distinction between innate and acquired processes, though Chapter Nine does not discuss it in this light, in that it assumes that some processes reflect the application of markedness conventions and thus presumably do not have to be learned, whereas other processes which do not mirror markedness conventions must be learned and therefore each feature in them carries a price tag, according to the evaluation measure of the theory. I do not have space here to discuss individual processes, but I think it must become fairly clear to anyone who investigates the distinction between innate and acquired processes that markedness theory drastically underestimates the number of processes which are innate.

This is most apparent in the case of processes that govern substitutions, as opposed to those that govern only lexical representation. Substitution-processes are those that may apply across morpheme boundaries. Whether morpheme boundaries block the application of a given process or not is a language-particular matter. But markedness conventions are language-universal, and thus it follows that either all of them are potentially applicable across morpheme boundaries or none of them are. Since the latter alternative would deny that *any* substitutions are natural, Chapter Nine must make all markedness conventions potentially applicable across morpheme boundaries. Then it is necessary to put severe constraints on the application of the conventions as substitutions, because if they applied freely they would reduce all phonological representations to [pa] or some such verbal pabulum. In my view innate processes ARE in fact freely applicable at the onset of language acquisition and do reduce the child’s mental representations of the words he knows to this or similar varieties of pabulum. His acquisition of pronunciation, accordingly, involves successively suppressing, limiting, or ordering those innate processes which stand between him and a mature pronunciation. The processes that survive acquisition thus constitute the natural core of the phonological system underlying his speech.

The conventions of markedness theory, on the other hand, cannot be suppressed, limited or ordered. Therefore Chapter Nine must seek other ways to constrain their application. What is proposed is Linking, which is basically the condition that a markedness convention may apply only to a substitute produced by the immediately



preceding application of a rule or of another markedness convention. The constraint is certainly a strong one, and it embodies at least the following claims. First, in limiting linking to apply only to substitutes, it claims that processes which apply in the context of substitutes (as opposed to those which apply to substitutes) are acquired processes. This seems impossible to support. Second, a process is claimed to be acquired unless it applies in immediate feeding order. Actually there are excellent examples of natural processes applying in nonimmediate feeding, antifeeding, bleeding, antibleeding, and even no order in particular. In fact, to accept this claim would be to renounce the insight that processes can be reordered in linguistic change, since it would not be a single process under analysis. Third, linking, in claiming that a markedness convention can apply only to substitutes, claims that a natural process discriminates between the derivational histories of segments to which it would apply. Chapter Nine cites some examples from Slavic as evidence, but these are refuted in a forthcoming paper by Bill Darden, and there seems little point in pursuing the claim further, except to ask why, if a natural process is one explicable in terms of phonetic limitations on our speech capacity, it should be more sensitive to derivational history than phonetic identities.

Substitutions are quite commonly limited by boundaries, or are limited to subsets of the natural classes involved in the most general form of a process, or are inapplicable in careful speech (i.e. are 'optional'). The flap process of English, a classic example of a natural process, exhibits all three types of limitation. Yet since markedness conventions cannot be limited language-specifically, we have a fourth claim: that processes with such limitations (not, notice, merely the limitations) must be acquired. 'If absurdities are not to result', it is necessary to exclude markedness conventions exhibiting sequential constraints from application as substitutions (343, note 22). This suggests that all assimilative substitutions must be acquired! Sixth, it is stated that a markedness convention applies to all the substitutes output by a rule, or to none. This condition seems to be required by Chapter Nine's device for blocking the application of otherwise applicable conventions, since the device does not allow for blocking SOME substitutions, and therefore conventions must apply to all or none. There is an argument that the reason German front-round vowels do not become unround by linking is that the rule that produces them, umlaut, also produced a vowel [e] which was already unround (432f). What this argument proposes, then, is that a natural process (unlike the rain of the Sanskrit grammarians, which fell on full and empty pails alike) may not apply at all if in any instance it applies vacuously. Pāṇini's revenge comes on the last page of SPE, however, where in a discussion of how markedness conventions apply to the output of the English vowel shift, a single application is said to change  $\bar{o}$  to  $\bar{a}$  and also (in an obvious allusion to the last rule of the Aṣṭādhyāyī)  $\bar{a}$  to  $\bar{a}$ .

To understand why Chapter Nine was led to such bizarre claims about substitution we have had to refer repeatedly to its assumption that natural processes reflect metagrammatical conventions, which are immutable and therefore subject to no



language-particular control except an all-or-none blocking device. There are, I think, much better ways to use markedness in a theory of substitution. Chapter Nine does not, in fact, use marks at all in substitutions, though most markedness adherents would take the view that every natural substitution unmarks the features to which it applies, in the sense that it supplies a value which is expected or natural for one which, in the context in question at least, is unnatural. To revise the substitution theory appropriately would entail that the language-particular grammatical apparatus of the theory refer directly to which marks are lost when. In fact, to reflect conditions of the generality of processes (for example, that if [y] is unrounded so is [ø]) it must refer not to which marks are lost when, but to which conventions apply when, and how generally, and furthermore under what stylistic conditions. At this point what began as a revision of the conventionalist view ends as a *reductio ad absurdum* of it: since the conventions are subject to exactly the sort of language-particular restrictions as natural processes, the distinction between conventions and processes becomes empty.

The other half of the theory concerns lexical representations and thus processes whose function is to restrict representations rather than to cause substitutions in them. Since markedness conventions are process-like, in that they imply substitutions, as opposed to the 'redundancy conditions' of Chapter Eight, Chapter Nine takes a laudable though apparently unwitting plunge into loan phonology. For what appears to be merely a condition on underlying representations often shows itself to be a process in the nativization of loan words, where confronted with an admissible segment, the 'condition' supplies a specific change. Research on the acquisition of phonology likewise suggests that conditions on underlying representation in a language originate as innate processes which, in child speech, often apply near the surface to cause actual substitutions, and then (instead of being suppressed) are ordered so that they are manifest only in restrictions on underlying representation. Furthermore, in relaxed and allegro speech, certain constraints on underlying representation appear in low-level substitutions, through unordering.

For example, children's pronunciation of English often exhibits an innate process deleting glottal stops, so that [bʌʔŋ] 'button' becomes [bʌŋ]. When these deleted glottals at last appear in the child's speech we might assume that the child has suppressed the deletion process. In fact it appears that he has merely ordered it to the north of the process which substitutes glottal for [t], and that it is this deletion process (not an unrelated, coincidental condition) which accounts for the inadmissibility and unpronounceability of nonderived glottal stops in English. It is no accident, therefore, that adult relaxed speech commonly exhibits deletion (via unordering) of derived glottal stops. Nor is it an accident that in foreign words with glottal stops which cannot be analyzed in the English system as derived from [t] are deleted.

While examples like this, which can readily be multiplied, support the view that restrictions on underlying representation are process-like, they also cast doubt on



the previous view that they are redundancy rules. Generative phonology would have treated the absence of the glottal stop in English underlying forms as due to a redundancy rule such as

$$\begin{bmatrix} -\text{Cns} \\ -\text{Voc} \\ -\text{Voi} \end{bmatrix} \rightarrow [+ \text{Cnt}]$$

which implies, counterfactually, that a glottal in a foreign word would become [h]. And if the previous theory rejected such implications, it would give up relating restrictions on underlying representations to the substitutions found in loan phonology. This would be tantamount to claiming that how we make foreign words admissible has nothing to do with why they are inadmissible.

All this is by way of prefacing my claim that markedness conventions are nothing but natural processes. The good thing about Chapter Nine is that in accepting conventions it abandons redundancy rules and conditions, which are even less process-like than conventions. All that remains to be abandoned then is the view that conventions are not processes.

First, however, a digression to consider Chapter Nine's conclusion that markedness furnishes no basis for a theory of admissibility. Since *Kindersprache*, it has seemed clear that, in some sense, the admissibility of the marked value of a feature implies the admissibility of the unmarked value, and it is therefore particularly puzzling that Chapter Nine sees no relation between markedness representation and admissibility.<sup>1</sup> The admissibility theory Chapter Nine proposes has nothing to do with markedness representations. It defines the degree of inadmissibility of a form as the reciprocal of the number of features required by a rule which 'distinguishes' the form from the rest of the forms of the lexicon. This is immediately suspicious because it denies the intuition that, whereas inadmissible forms are indeed inadmissible to various degrees, any admissible form is as admissible as any other. Furthermore, it emerges on closer inspection that the new proposal is viciously circular, for it requires us to know which forms are admissible in order to determine which are inadmissible (and how inadmissible). Consider a case of two forms that share a single putatively inadmissible property (e.g. *husband*, *asbestos*, cited [150, note 105] as exceptions to

<sup>1</sup> Chapter Nine abandons the redundancy theory of admissibility of generative phonology without showing why it had to be abandoned, though it does recognize that there had been some 'idle controversy' (418, note 7) over the previous theory. However, the idle controversy in question was not, as Chapter Nine claims, over the fact that the theory incorporated an evaluation measure, but rather over the fact that the measure, which rejected a redundancy rule if it supplied less features to the lexicon than were required to state it, claimed in effect that whether one could say a word with, for example, initial *ft* depended on how many forms he knew with initial *st* (sic). This is because the acceptance of a rule defining initial consonants before *t* as nongrave depends, according to the theory, on how many nongrave specifications it would eliminate from the lexicon, and hence on how many forms the lexicon has with initial *st*.

It had also been idly asked how the fact that a feature value was redundant could explain why the opposite value should be hard to pronounce.



the usual voicelessness of obstruent clusters in English). Neither can be distinguished from the lexicon by a rule mentioning their voiced clusters so long as the other is in the lexicon. The theory makes no allowances for rules distinguishing CLASSES of forms from the lexicon, and if it did it would require us to know whether a class or its complement is the admissible one: otherwise we would find forms with voiceless clusters just as inadmissible as those with voiced clusters. An auxiliary hypothesis that, for example, made the minority class the inadmissible one would lead to such absurdities as that /y/ is inadmissible because there are less of them than of /w/s in the lexicon.

In the absence of an admissibility theory, Chapter Nine fails totally as a theory of lexical representation. If lexical processes – which may apply or not in a language, depending on its underlying oppositions – are replaced by convention-interpreted marks in the representations of lexical items, there is no distinction between admissible and inadmissible segments at all. In a five-vowel language like Latin markedness representations define /o/ as one degree more complex than /u/, but since /ü/ and /i/ are also only one degree more complex than u (409), a word like /hümo/ or /himo/ is defined as just as Latin as /homo/. This cannot be repaired by reweighting the complexities of these vowels, because there are other languages which admit /ü/ or /i/ but not /o/. In fact, the theory implies that a language with a dozen or more vowels of which /a/ is highly frequent is no less expected than one with five vowels in which more complex vowels like /e o/ are fairly frequent, since the total marks in their lexicons might be equal. That is, the theory fails to account for the fact that languages have specific limited phonological inventories.

This brings us to the question of whether marks themselves have any empirical reflexes. I have pointed out that Chapter Nine finds no use for them in interpreting substitutions or even in determining the admissibility of underlying representations. The only role markedness representations play in Chapter Nine, in fact, is in the representation of the lexicon. One must question, however, even this role. There is an argument in Chapter Nine (419) that speakers identify all nasals before stops, which in markedness representation are simply nasals unmarked for point of articulation, as /n/, this being the least marked nasal.<sup>2</sup> In fact, what evidence we have regarding this suggests that while the nasals of *lint* and *link* are identified as /n/ the nasal of *limp* is identified as /m/. The reason for this, from the viewpoint of natural phonology, is that the natural process making nasals alveolar is limited in English to nonlabial nasals (otherwise we could not pronounce words like *mouse*) although it does (as the pronunciation of foreign names like *Nguyen* with [n] indicates) still apply to [ŋ]s which can't be derived from /ng/. The limited process thus bars /ŋ/ from lexical representation, but not /m/, and therefore *limp* is permitted, unlike *link*, to be represented with its surface nasal. If this account is correct, and there are many

<sup>2</sup> The claim (419, note 10) that this explains why *sink* is spelled with *n* rather than *m* must be rejected. The reason is obviously that in English assimilation to a following stop is limited to /n/ and does not apply to /m/ in the first place.



parallel examples, then markedness theory actually makes a false claim about the nature of underlying representation, and, since there is no provision for limiting markedness conventions in the language-particular manner indicated for English, the theory is stuck with its claim.

And in general it seems clear that markedness representation actually plays no role whatever in our identifications of underlying segments. The least marked consonant in initial preconsonantal position (416) is /s/, for example, while the least marked consonant elsewhere is /p/ or /t/. If the notion 'least marked consonant' had any independent reality – i.e. if it referred to anything more than the resultant segment after all natural processes applicable to consonants have applied – then we should expect to find some identification of /s/ with /p/ or /t/ in the contexts indicated. For example, we might expect a child who inserts schwa between /s/ and a following consonant, thereby removing /s/ from preconsonantal position, to change the /s/ to the appropriate 'least marked consonant' /p/ or /t/. But of course no such thing happens.

The simplest explanation for why markedness representations play no role in phonological systems would be that they do not exist, and thus that markedness conventions do not exist either. There are two important arguments for this conclusion. First, as McCawley has pointed out to me, marks are a measure of the unnaturalness of a segment, on one hand, and on the other, a sort of exception feature that renders a markedness convention inapplicable. But the implicit and inescapable suggestion of this view, that the degree of unnaturalness of a segment is a function of the number of natural processes to which it is subject, is clearly false. The schwa vowel, for example, can change to virtually any other primary vowel, and yet it is scarcely rare in the languages of the world: whereas the exceedingly rare [œ] seems to be subject to only two context-free processes, raising and fronting.

Second, if marks were real we would expect that the acquisition or loss of a mark in one segment would favor its acquisition or loss in another segment; thus given the following relations,

<i>mRound</i>	<i>uRound</i>
u	u
y	i
ɔ	a

we would expect the acquisition or loss of [u] to imply an acquisition or loss of [y] and [ɔ]. But this is not the case at all; there is no relation whatsoever between these segments although they are all identically marked for Roundness. The reason of course, is that there is no single natural process which unrounds any two of these segments. Contrast the case of [ø], which is always unrounded if [y] is, because it is a single process which unrounds them. (The opposite implication, that [y] must be unrounded if [ø] is, does not hold, such implications being unilateral.)

In fact it appears even that implicational laws, to the extent they hold, are nothing



more than empirical generalizations regarding the effects of natural processes. There is a natural process Affricate  $\rightarrow$  Stop which is the basis of Jakobson's implicational law Affricate implies Stop. But Affricates do not necessarily imply stops of course: there is another well-known process

$$\begin{bmatrix} \text{Pal} \\ \text{Stop} \end{bmatrix} \rightarrow \text{Affr}$$

and if it applies in a language the implicational law is violated. Jakobson was able to ignore these by assuming that, where there is no contrast, palatal affricates are phonemically palatal stops. Aside from the fact that such an assumption requires proof, we have to note that in the absence of a theory of natural processes the assumption also would suggest that we should find not only systems like

$$\begin{bmatrix} p & t & k \\ & \check{c} & \end{bmatrix} = /p t k k/$$

but also systems like

$$* \begin{bmatrix} & t & k & k \\ p f & & & \end{bmatrix} = /p t k k/.$$

The fact that the former system is possible and the latter impossible is a result of the fact that there is a natural process affricating palatal stops, but none affricating labial stops. What is crucial here is that we have two natural processes, Affrication and Deaffrication, which happen to conflict in the case of palatals. In natural phonology such conflicts are resolved merely by suppressing or ordering the processes. In markedness theory this is impossible. It cannot represent either the stops *or* the affricatives as marked or unmarked. Nor can it merely withhold markedness representation from them, as Chapter Nine proposed for certain cases, because unlike the cases described in Chapter Nine,  $\check{c}$  becomes  $k$  and vice versa.

The conclusion that must be drawn from such facts, and they can readily be multiplied in this and other forms, is that marks, and markedness conventions, are mere appearances, and that what underlies the impression of reality they bear is, in fact, the innate system of natural processes.

Ohio State University

#### REFERENCES

- Chomsky, Noam, and Morris Halle  
1968 *The Sound Pattern of English*.  
Jakobson, Roman  
1968 *Child Language, Aphasia, and Phonological Universals*.  
Stampe, David  
1969 *The acquisition of Phonetic Representation. Papers from the Fifth Regional Meeting of the Chicago Linguistic Society*.