
MARKEDNESS*

MORRIS HALLE**

The questions that concern me in this paper are the direct outgrowth of the study of English phonology that N. Chomsky and I have been engaged in for about a decade.¹ This investigation has brought out certain inadequacies in the theoretical framework, and I shall discuss here a proposal that we have made in order to remedy these inadequacies. It is worth remarking that the inadequacies in our theoretical framework were discovered only because we were trying to push the theoretical machinery to its limits, to subject it to the linguist's equivalent of the engineer's break-down test. Such a test, however, can only be conducted by a detailed examination of the functioning of the theoretical apparatus. This in turn pre-supposes that, on the one hand, there be a high degree of clarity as to how a given body of data is to be characterized formally by the descriptive machinery and that, on the other hand, these characterizations, i.e., the phonological rules and their formal properties, be taken fully as seriously as the linguistic facts that they characterize. In the following discussion the phonological rules, therefore, occupy the central position. They are not just more or less accidental formulations, of interest to the linguist only by virtue of the facts that they reflect. They are, rather, the essential organon, the main tool for acquiring deeper knowledge about the phenomena in question.

It is all but self-evident that in every science there are more and less effective ways of describing particular observations. Thus, all of us would normally prefer a simple formula to a list even if the formula and the list cover precisely the same body of facts. Karl Verner is justly esteemed as one of the leading linguists not because he discovered any new facts, any new Germanic forms that were exceptions to Grimm's Law, but rather because he discovered the simple formula —Verner's Law—that succinctly characterizes the list of exceptions. Since different formulations of a given body of data are thus clearly not all regarded as equivalent, it is essential to attempt to discover the general principles on the basis of which one formulation

* This work was supported in part by the U.S. Air Force [ESD Contract AF 19 (628)—2487] and the National Institutes of Health (Grant MH-13390-01).

** Massachusetts Institute of Technology.

¹ The present study summarizes the main points of the final chapter of Chomsky and Halle (1968).

is to be valued over another. As is well known, we have proposed that an appropriate evaluation measure for phonological rules as well as for lexical representations is the number of features required to state them: the lower this number, the greater the generality and the higher the value of the description.² We have supported this proposal by exhibiting certain cases where indeed it led to the correct results. As a very simple example we may cite the fact that (1a) is a more general, and hence more highly valued rule than (1b).

- (1) (a) $[i, e, \text{æ}] \rightarrow [\text{ü}, \text{ö}, \text{œ}]$
 $[-\text{back}] \rightarrow [+ \text{round}]$
- b) $[i, e] \rightarrow [\text{ü}, \text{ö}]$
 $[-\text{back}] \rightarrow [+ \text{round}]$
 $[-\text{low}] \rightarrow [+ \text{round}]$

There are, however, quite a number of instances where simple-minded feature counting fails to express the state of affairs properly. As examples consider (2):

- (2) (i) (a) $u \rightarrow i$ (b) $u \rightarrow \text{ü}$
 $\begin{bmatrix} +\text{high} \\ +\text{back} \\ +\text{round} \end{bmatrix} \rightarrow \begin{bmatrix} -\text{back} \\ -\text{round} \end{bmatrix}$ $\begin{bmatrix} +\text{high} \\ +\text{back} \\ +\text{round} \end{bmatrix} \rightarrow [-\text{back}]$
- (ii) (a) $k \rightarrow \text{č}$ (b) $k \rightarrow p$
- (iii) (a) $k \rightarrow \text{č}$ (b) $\text{č} \rightarrow k$

In (2i) the left rule requires more features to state than the right hand rule but it is by no means obvious that there is a real distinction in the generality of the two rules. In (2ii) $k \rightarrow \text{č}$ requires more features than $k \rightarrow p$, but $k \rightarrow \text{č}$ is certainly a more general rule than $k \rightarrow p$. Finally the rule $k \rightarrow \text{č}$ is much more common than the rule $\text{č} \rightarrow k$ but this fact finds no reflex in the number of features that are required to formulate these two rules. Thus, we have here a number of instances where feature counting clearly will not make the sort of choices that an evaluation measure would be expected to perform.

Feature counting fails yet in another way. We have proposed (Halle [1959] and elsewhere) that the regularities in the phonological composition of morphemes in a language would be properly expressed if lexical items were represented by the minimum number of features that allowed the item to be unambiguously identified in the lexicon. The omitted features would then be specified by general rules (morpheme structure rules). As a consequence, in the lexical representations of a language that had only the three vowels /a i u/ these vowels would be specified by fewer features than in a language that had the five vowels /a i u e o/, as shown in (3).

² For a detailed discussion of the problems of the evaluation procedure see Stanley (1967).

(3) (a)		a	i	u
high	-	+	+	
back		-	+	

(b)		a	i	u	e	o
high	-	+	+	-	-	
back		-	+	-	+	
low	+			-	-	

This is, of course, the desired result since in the language (3a), which has fewer vowels, more features are specified by general rule than in (3b), which has more vowels. In the three-vowel language (3a) the feature *low* is determined by the rule that high vowels are nonlow, and nonhigh vowels are low; the feature *low* cannot be determined by rule in the five-vowel language (3b) and hence cannot be omitted in the lexical representations there without leading to confusion among distinct items.

A parallel argument holds for the representations of the vowels in (4),

(4) (a)		æ	ü	i
high	-	+	+	
back		-	+	

(b)		æ	ü	i	ö	γ
high	-	+	+	-	-	
back		-	+	-	+	
low	+			-	-	

with the result that the two vowel systems in (4) are represented by the same set of features as the vowel systems in (3). The only difference between the vowel systems in (3) and (4) is the respective rules specifying the features that have been omitted in the lexical representation. This minor formal difference, however, reflects inadequately the very profound differences that must exist among languages with vowel systems such as those in (3) and those with vowel systems such as (4).

The examples of the failure of the proposed evaluation criterion raise seriously the question whether it is correct to weight all features equally. There appear to be clear cases where some features are more equal than others. Or less facetiously, there are grounds for questioning the completely formal, abstract view of features that has been fundamental to our approach, and that led us, as a matter of principle almost, to neglect the intrinsic content of the features and the rules. The question that posed itself, therefore, was how to take account of this intrinsic content of features and rules.

The direction towards a solution became clearer to us as we attempted to resolve another difficulty that stemmed from our decision to omit features in lexical representations. Lightner (1963) showed that when features were omitted in lexical representations, it was possible to effect certain spurious economies by appropriate

convention (5vb) replaces [u round] by [—round] yielding the feature complex

$$\begin{bmatrix} +\text{low} \\ -\text{high} \\ -\text{back} \\ -\text{round} \end{bmatrix}$$

or [æ]. In (6a) we have given the feature markings of the main vowel types, and in (6b), their phonological feature composition which will be derived from the markings in (6a) in the fashion just described.

(6) (a)

	a	i	u	æ	e	o	ɔ	ü	ĩ	ö	ə	æ
low	u	u	u	m	u	u	m	u	u	u	u	m
high	u	u	u	u	m	m	u	u	u	m	m	u
back	u	—	+	m	—	+	u	—	+	—	+	m
round	u	u	u	u	u	u	m	m	m	m	m	m

(b)

	a	i	u	æ	e	o	ɔ	ü	ĩ	ö	ə	æ
low	+	—	—	+	—	—	+	—	—	—	—	+
high	—	+	+	—	—	—	—	+	+	—	—	—
back	+	—	+	—	—	+	+	—	+	—	+	—
round	—	—	+	—	—	+	+	+	—	+	—	+

We can now return to the problems that arose as a result of our decision to omit predictable features in the lexical representations of morphemes. We recall that one of our reasons for omitting these features was that we wished to extend our evaluation criterion, which consists in counting features in rules, to the lexical representations. We now have another way of evaluating lexical representations. We assume that instead of being represented by matrices such as (6b) in which + and — figure, lexical items are represented by matrices such as (6a) in which features are marked or unmarked. We now postulate that in assessing the cost of a given lexical representation we count the number of feature entries that are not unmarked. Thus, for example, given the matrix (6a) the vowel [æ] has a cost of two features, the vowel [ö] a cost of three and the vowel [a], a cost of zero.

It is to be noted that when segments are represented with feature markings there is no natural way in which one may omit to specify a feature. Thus, in (6a) there are only marked or unmarked, but no omitted features. This fact has bearing on the

solution of the difficulties discovered by Lightner. If lexical representations are framed in terms of marked and unmarked features, we need not have any unspecified features in the lexicon. We recall that the primary motivation for omitting features in the lexicon was that this served as an evaluation measure. We have now an evaluation measure that does not require us to have features unspecified in the lexical representations. We can, therefore, dispense with the procedure of leaving features unspecified in the lexicon, thereby avoiding the serious problems to which Lightner first drew attention.

It may finally be noted that there is an alternative way of looking at the tables in (6). Rather than suppose that the lexical representations are framed in terms of feature markings to be translated by the conventions (5) into plusses and minuses of phonological features, we may suppose that the lexical representations are framed in terms of plusses and minuses and that the function of the conventions (5) is to provide us with a means of assessing the cost of a given configuration of plusses and minuses. We do this by discovering the particular configuration of feature markings that is translated into the appropriate plusses and minuses by the conventions (5) and then counting the number of marked features in the former. It is important that there is no substantive difference between these two procedures; they are notational variants of one and the same thing.

We must now return to the problems that we raised in connection with the vowel systems in (3) and (4). Why is the set [a i u] preferable to the set [æ i ü]? The answer is obvious if we examine the table (6a). The former set has 2 marked features in toto; the latter has 6 marked features. Thus in order to make the proper choice in these cases, we must require formally that in choosing a vowel system, the one least complex in terms of markedness be chosen. Observe that if this is made a formal requirement, then in the lexical representations of a language that has a 3-vowel system, the vowels will always be [a i u]. If the phonological rules require for their input the system [æ i ü] this will be achieved by a redundancy rule that will operate on the lexicon. The difference, therefore, between a language with the system [a i u] and one with [æ i ü] will be that a redundancy rule will be required only in the latter case, but not in the former.

It is, however, readily apparent that complexity as defined here will not provide an answer in all cases. Compare the 5-vowel systems in (7).

(7)

i	u	i u i ü	i u	i u	i ü u
e	o			e	e
a		a	æ a ɔ	a ɔ	a

It is clear that the first is to be preferred over the rest. To formally implement this choice we need two further principles.

(8) Other things being equal, a system in which more features have only the entry u is preferable to a system in which fewer features have only the entry u.

This principle immediately eliminates all but the first two systems in (7). The meaning of (8) is that it prefers the system that makes maximal utilization of each feature for purposes of marking. It, therefore, builds into the theory a bias in favor of *more symmetry*. Observe that the symmetry is of a very abstract sort.

The choice between the first two systems in (7) needs yet another principle. We believe that this principle is that of feature hierarchy. Since we have been unable to discover a complete feature hierarchy, we shall give only a small part of it here as (9).

(9) No vowel system can have a segment marked for round, unless it also has a segment marked for high.

The question of feature hierarchy is certainly of the greatest importance but it has so far resisted all of our attempts to find a proper solution. We see the difficulty immediately when we consider 4-vowel systems. Given the principles enunciated so far, a 4-vowel system must include [i u a] plus a vowel of complexity 2 that is not marked for rounding. We see in (6a) that there are 3 candidates that satisfy these conditions [æ e o]. It would seem that a proper extension of the feature hierarchy should provide the correct choice but so far we have been unable to discover how to do that.

In spite of the present unsatisfactory situation I think that it is reasonably clear that marking conventions like those in (5) supplemented by certain further principles will enable us to overcome the inadequacy noted in connection with the existence of different vowel systems. It should be underlined that by means of the marking conventions we are taking specific account of the content of the features, for each feature is interpreted by a particular convention in some particular context, in its own specific way.

We must now turn to the second set of examples that we discussed at the beginning, namely the examples illustrating a certain inadequacy in the formulation of rules, given the present theoretical framework. Consider first the examples in (2i). The reason that (2ia) is more complex than (2ib) is that we have no way in the present theory to cause one feature to change in tandem with a change in some other feature. It seems in many instances, however, e.g., in (2i), that this is precisely what happens, when backness changes in many cases it seems normal for rounding also to change. Our problem, therefore, is to discover some device that would allow for such changes in tandem.

It may not be obvious that such a device is already at hand in the marking conventions. In particular, convention (5va) says that in nonlow vowels rounding must agree with backness. Our problem, therefore, is now to bring to bear the marking conventions on the interpretation of individual phonological rules. This can be done in the following way. We can require that whenever a phonological rule changes a feature in such a way that the segment undergoing the change would satisfy the environment of any marking convention, then the segment will also be provided with the feature specification that the convention provides for the unmarked feature.

Thus, a rule

$$[+high] \rightarrow [-back]$$

will automatically also imply [-round] by virtue of convention (5va). However a rule such as

$$[+low] \rightarrow [-back]$$

will not link to any convention since convention (5va) applies only to [-low], and not to [+low] vowels. If the proposed procedure is adopted in general, it will now be rather more difficult to state (2ib), for we shall have to have some way of undoing the effects of the convention (5va). The appropriate procedure would be to add a second line on the right hand side of the arrow so that the rule would read:

$$[+high] \rightarrow \left[\begin{array}{l} -back \\ +round \end{array} \right]$$

The number of features that have to be mentioned in a rule does not in all cases correspond to the number of features that are actually changed by the rule. Thus under the conventions just discussed the rule

$$[+high] \rightarrow [-back]$$

changes both rounding and backness; whereas the more complicated rule

$$[+high] \rightarrow \left[\begin{array}{l} -back \\ +round \end{array} \right]$$

changes only backness and leaves rounding intact.

An even more complicated example is that involved in the Great English Vowel Shift. The Vowel Shift accounts for such alternations as

divine	divinity	profound	profundity
keep	kept	lose	lost
profane	profanity	verbose	verbosity
i ~ ī	→ āy	ɔ ← u ~ ū	→ āw
e ~ ē	→ īy	ɔ ← o ~ ō	→ ūw
æ ~ ǣ	→ ēy	a ← ɔ ~ ɔ̄	→ ōw

In *The Sound Pattern of English* (Chomsky and Halle, 1968) arguments are presented showing that these alternations are actually alternations between tense and lax vowels and that the diphthongization and changes in vowel quality in the tense vowels are due to the operation of special rules, in particular those in (10).

$$(10) (a) \varnothing \rightarrow \left[\begin{array}{l} \text{Glide} \\ \alpha\text{back} \end{array} \right] / \left[\begin{array}{l} V \\ +\text{tense} \\ \alpha\text{back} \end{array} \right] \text{---}$$

$$(b) (i) \left. \begin{array}{l} \left[\begin{array}{l} \gamma\text{high} \\ -\text{low} \end{array} \right] \rightarrow [-\gamma\text{high}] \\ \left[\begin{array}{l} \beta\text{low} \\ -\text{high} \end{array} \right] \rightarrow [-\beta\text{low}] \end{array} \right\} / \left[\begin{array}{l} +\text{tense} \\ \alpha\text{back} \\ \alpha\text{round} \\ V \end{array} \right]$$

The vowel shift rule (10b) abbreviates 2 pairs of disjunctive rules as shown in (11).

- (11) (ai) $\begin{bmatrix} +\text{high} \\ -\text{low} \end{bmatrix} \rightarrow [-\text{high}]$ (aia) $\begin{bmatrix} -\text{high} \\ -\text{low} \end{bmatrix} \rightarrow [+high]$
 (bi) $\begin{bmatrix} +\text{low} \\ -\text{high} \end{bmatrix} \rightarrow [-\text{low}]$ (bia) $\begin{bmatrix} -\text{low} \\ -\text{high} \end{bmatrix} \rightarrow [+low]$

The first three of these rules do not link to any of the marking conventions, hence they have the following effects:

$$\begin{array}{lll} \bar{i}y \rightarrow \bar{e}y & \bar{e}y \rightarrow \bar{i}y & \bar{\alpha}y \rightarrow \bar{e}y \\ \bar{u}w \rightarrow \bar{o}w & \bar{o}w \rightarrow \bar{u}w & \bar{\alpha}w \rightarrow \bar{o}w \end{array}$$

Since (11ai) and (11aia) as well as (11bi) and (11bia) are disjunctive, (11aia) does not affect segments that have undergone (11bi). However, segments that have undergone (11ai) or (11aia) are not exempt from (11bi) or (11bia). Hence, (11bia) applies to (\bar{e} , \bar{o}) which are the reflexes of (\bar{i} , \bar{u}) after having undergone (11ai). These now become [+low]. But this change links these vowels to marking conventions (5iv) and (5vb) with the results shown here.

$$\begin{array}{lll} 11bia & 5iv & 5vb \\ \bar{e}y \rightarrow \bar{\alpha}y \rightarrow \bar{a}y \rightarrow \bar{a}y & & \\ 11bia & 5iv & 5vb \\ \bar{o}w \rightarrow \bar{\alpha}w \rightarrow \bar{a}w \rightarrow \bar{a}w & & \end{array}$$

In other words, the marking conventions state that when \bar{e} , \bar{o} become low it is more natural for them to become simultaneously also nonround and back, (i.e., / \bar{a} /) rather than preserve their original rounding and backness, and the facts of the Great English Vowel Shift as well as analogous facts from other Germanic languages seem to support this empirical claim implicit in the marking conventions.

Convention (5va) illustrates a further interesting property of the marking conventions. Consider a language like Finnish in which there is vowel harmony with respect to the feature *back*. That is, in this language a word will contain only one of the two sets of vowels shown in (12a);

$$(12) (a) \frac{i \ e \ u \ o \ a}{i \ e \ \bar{u} \ \bar{o} \ \bar{\alpha}}$$

which can be accounted for by two rules such as (12b):

$$(12) (b) (i) \ V \rightarrow [\alpha\text{back}] / \#C \begin{bmatrix} V \\ \alpha\text{back} \end{bmatrix} X \text{---}$$

$$(ii) \begin{bmatrix} V \\ -\text{low} \\ -\text{round} \end{bmatrix} \rightarrow [-\text{back}]$$

Rule (12b) is an abbreviation of two rules:

$$\begin{array}{l} V \rightarrow [+back] \\ V \rightarrow [-back] \end{array}$$

and hence affects [u] turning it into [ü]. The question now arises whether or not convention (5va) should apply here. Clearly the answer is *no*, for if the convention were to apply, we should have to give the rule in quite complex form. Instead of (12) we should have to have

$$(13) \ V \rightarrow \begin{bmatrix} \alpha\text{back} \\ \beta\text{round} \end{bmatrix} / \#C \begin{bmatrix} V \\ \alpha\text{back} \end{bmatrix} X \begin{bmatrix} \beta\text{round} \end{bmatrix}$$

We observe, however, a further fact: the marking convention (5va) does not apply to all cases to which rule (12bi) applies; rule (12bi) applies to low vowels as well as to nonlow vowels. This observation suggests that the way to avoid the difficulty just discussed is by imposing the constraint (14):

- (14) A marking convention applies either to all or to none of the segments affected by a rule.

Since vowel harmony affects all vowels, whereas (5va) affects only non-low vowels, convention (5va) does not link to the vowel harmony rule, which can therefore be given in the simple form of (12bi).

We have found that (14) correctly handles a number of other cases as well: it provides the correct answer in the various types of umlaut rules, and it also accounts for the appearance of (θ) in place of the less marked (s) in cases such as Grimm's Law and the Semitic spirantization. These are extremely strong bits of evidence supporting the proposed constraint.

The proposed marking and linking conventions are, thus, supported in two distinct ways. On the one hand, they are justified by the fact that the predominant types of phonological system exhibit the properties imputed to them by the proposed conventions. The same conventions appear, furthermore, to be justified by the fact that they account for the predominant types of phonological rules incorporated into the grammar of each individual language. We have here two disjoint sets of facts that appear to be accounted for by a single theoretical extension, the introduction of the markedness conventions. That two disjoint sets of facts appear to be accounted for by the same extensions of the theory, i.e., the introduction of the markedness conventions, is hardly an accident. It suggests that further explorations in this direction are likely not to be totally fruitless.

BIBLIOGRAPHY

- Halle, Morris, *The Sound Pattern of Russian* (The Hague, 1959).
 — with Noam Chomsky, *The Sound Pattern of English*, Harper and Row, Inc., New York, 1968.
 Lightner, Theodore, "A Note on the Formulation of Phonological Rules", *Quarterly Progress Report of the Research Laboratory of Electronics, M.I.T., Cambridge, Mass. No. 68*, January 15, 1963.
 Stanley, Richard, "Redundancy Rules in Phonology", *Language*, 43, 393—436 (1967).
 Trubetzkoy, N. S., *Principles de phonologie* (translated by J. Cantineau), Paris, 1949.

Akhmanova:

The misunderstanding that is bound to arise here is due to faulty terminology: what is proposed here is *not* phonology; it is *morphology*. As such it is very useful and, clearly one of the possible approaches to the problem (particularly acceptable to those who are looking for "universals", for the features that are, supposedly, common to all mankind), But it is completely misleading to call it "phonology". The task of phonology is to discover and determine which of the innumerable possibilities of sound production (the sounds the human organs of speech can produce) are used in different languages as "diacritics" of the "feature level"; in other words, investigate those contrasts and oppositions which are perceived and used by the speakers of different languages.

Martens:

Anmerkung zum Diskussionsbeitrag von Herrn Wiesinger: Es ist methodisch nicht glücklich, "Vokal e" den "normalsten" deutschen Vokal zu nennen (auf Grund der statistischen Frequenz). Eine solche Klassifizierung läßt die Gruppierung nach Akzentuierungsdistribution unberücksichtigt. Man sollte nicht alle (ə), die *nur* in *nicht-akzentuierter* Position auftauchen, zusammenwerfen mit [e], das *nur* in *akzentuierter* Position vorkommt.

Martinet:

What advantage do we derive from such a complicated scheme as the one which has just been presented? I would think that the reason why the vocalic pattern with *i—u* as the highest vowels should be more normal or frequent than the ones with *u* and *ü*, is that *i—u* secure maximal differentiation through an opposition of minimal and maximal front cavities, whereas *u* and *ü* require both medium cavities. It seems that we have here a conflict between those who think that all languages are basically identical and that the duty of linguists is to discover this identity, and those who think that this duty lies in determining what makes a language different from all others within the framework of what is recognized as necessary for an object to be called a language.

Wurzel:

I. Was wollen wir mit der Theorie der Markiertheit?

1. Eine Theorie phonologischer Universalien (Aufbau von phonologischen Systemen) oder
2. Möglichst einfache Lexikoneintragungen und Regeln in *einer* bestimmten Sprache. Beides kann sich sehr stark widersprechen (Beispiel: Phonem [e] im Deutschen häufigster und normalster Vokal).

Wenn wir uns für 2. entscheiden, dann gibt es auch notwendigerweise Markierungswechsel in der Sprachgeschichte.

II. Problem der phonologischen „Nichtakzeptierung“ des germ. Umlauts (das nicht finn.-ugr. oder türk. vorkommt) kann durch Markierung vielleicht gut erklärt werden. Verbreitung der Entrundung gibt es in nahezu allen germ. Sprachen. Gerundete *vordere* Vokale entwickeln sich sekundär als [m rund].

Streben nach Beseitigung der Markierung.

ü, ö sind $\begin{bmatrix} u \text{ dunkel} \\ m \text{ rund} \end{bmatrix}$, Entrundung geschieht nicht durch Wiederzusammenfall mit *u, o*, (denn wie gesagt *vordere* Vokale sind im Deutschen offensichtlich „normaler“. Es wird das

markierte Feature verändert: $\begin{Bmatrix} \ddot{u} \\ \ddot{o} \end{Bmatrix} \rightarrow \begin{Bmatrix} i \\ e \end{Bmatrix}$

i, e haben die features $\begin{bmatrix} u \text{ dunkel} \\ u \text{ rund} \end{bmatrix}$