To appear in The Last Phonological Rule: Reflections on Constraints and Derivations, edited by John Goldsmith. University of Chicago Press, 1993. © by The University of Chicago. All rights reserved. For review and analysis only -- not for publication.
important insights at the core of lexical phonology that seem to easily get lost—such as the thorough-going finiteness of morpheme-structure conditions and lexical phonological rules. The only major criticism that the present paper aims toward is that all phonological rules which apply at a particular level have the explicit function of moving a representation as far as possible toward instantiating the phonotactics of that level; that these rules, within a level, are not ordered, that rules which apply across levels do not necessarily have such properties, in general, but that these rules do not give rise to derivations (i.e., to derivations with intermediate stages); that the levels of a phonological account are few in number, and that their properties are largely independent of one another (pair suggestions of structure preservation); and, finally, trisegmentalism: that greater attention to what defines well-formedness at a given level will lead to a far simpler overall grammar. The present paper is perhaps no more than a proprietorial to a proper and full treatment. 7

2.2. Representations, Levels, and Rules

All theories of phonology—and, more generally, of formal linguistics—can be usefully divided into theories of representations, of levels, and of rules. All these are potentially problematic notions, and the boundaries are often difficult to define. But this trisegmental division is nonetheless very useful, and worth the effort we expend on establishing it.

One of the notions, that of representation in the most familiar at present. Most of the work in phonology in the post-SPE period—1975–1990, let us say, following the publication of The Sound Pattern of English (SPE; Chomsky and Halle 1968)—has been in this domain. Over the last fifteen years, phonologists have taken it to be a matter of debate and exploration to find the most appropriate geometrical and algebraic models for representing phonological information. These explorations have included the development of abstract segmental structures and association lines, of metrical trees and grids, of feature geometries, of dependency relations among feature specifications, of “puzzles” approaches to segmental structure (involving hypotheses regarding the atoms that compose vowels and consonants), and so on.

It is worth bearing in mind that current openness to such discussion has not been achieved effortlessly. The generally unchallenged assumption throughout much of modern phonological thought has been that phonological representation was largely unrepresentable and consisted, in particular, of uninterpreted segments. The dominance of this position was reinforced by its centrality to both Bloomfieldian thought and to that of The Sound Pattern of English and the voices that were raised to question it had little impact on the global assumptions made by theorists of phonology during this period. Much of this has changed now, to be sure, but the change has been a means one. In the domain of rules and levels, the range of debated issues has been much less varied and much more restricted than in the domain of representations, but we may reasonably hope that what has been in large part a matter of focus on attention: the field, in the last fifteen years, has satisfactorily established for itself a class of adequate phonological models, it is now in a position to turn its attention to other, equally difficult matters whose consideration has largely, though by no means entirely, been put on hold during this period.

2.3. Levels

The notion of level is perhaps the single most important notion in modern linguistics, and there is a danger that our understanding of this notion may fade from our consciousness as various technical terms vie for our professional attention: the case could be made that linguists' appreciation of this notion has diminished as a result of certain competing interests. There is a sense in which we are all comfortable with the notion of levels in linguistic analysis, but I invoke the reader to consider some basic questions once again: Our first goal is to clarify the notion of level, and when we do so, we find, first, that it is broad (and quite well) explained in prescriptive writings; second, that generative grammar was originally conceived—no surprise!—as an attempt to questions formulated within this clear and traditional understanding of the notion of level, and third, that the notion of a level has passed from being, at its origination, a possible approach to the problem of linguistic levels, as a new of linguistic analysis which culminates in the way of a clear understanding of the notion of levels. 3

Like all important ideas, that of linguistic levels is very simple: a linguistic level is a way of looking at—or of describing—a linguistic expression. We may look at an expression from a syntactic perspective, and produce a syntactic level of representation; or from a morphological perspective, and produce a morphological level of representation, and so forth. In our usual linguistic way of thinking, there is an inherent ordering—or at least a relationship close to ordering—of these levels based on the relative size of the units that are established on each level of representation: if the units in the syntactic level (always, or typically) correspond to one unite units on the morphological level, then there may well seem to be an inherent ordering of these two levels, with one "above" the other, and, indeed, the...
term "level" might encourage such an addition to the concept. But this kind of wording is not adequate to the notion of level. Let us recall a simple example of how levels of analysis of an utterance may differ. A representation on the morphological level consists of units that are meanings and, as such, do not yield to spelling a sentence as, say, "Ago or [dog]", but rather as $\mu_g$, for example, to emphasize that it has no internal structural structure at this level. The analysis of an utterance such as the dog barked into morphophonemic level units, which may well include boundaries of (our theory allows there) meaning the morphemes, as in (1.3), does not constitute a representation on the morphological level—it is an analysis on a phonological or morphophonological level, as indicated previously by the kind of units (phonological units) that form the representation.

(1) a. The dog barked.
   b. [dog barked] [Phonemic Representation]
   c. Morphophonemic level

Levity, indeed, may be quite autonomous and independent of each other. The degree of autonomy of the various levels was a significant issue in the 1950s, Freeland (1961, for example), argued as though that the morphological level and the phonological (or, specifically, phonemic) level are autonomous and cannot be viewed as variables, i.e., a single linear hierarchy of levels. Pike's work greatly addressed this question, in the context of a theory involving five hierarchies of distinct levels: the phonological level, the morphological level, and the lexical level, as well as deep structure and surface structure. In other words, the level itself was defined in terms of a derivation. In a sense, this last sentence turns the logic of analysis on its head, in a way that echoes our own discussion. The most interesting example of this is found in A. Theories of Syntax (Chomsky 1956, 138), where the notion of deep structure is defined more simply as the result of lexical insertion into the output of the phrase structure rules. The picture we have in mind, a further condition is set there on the deep structure. A structure which is created by the phrase structure rules and to which the rules of lexical insertion essentially assign lexical elements is called a "generative phrase-marker." A generative phrase-marker is a deep structure only if, in addition, it is part of a derivation that includes a well-formed structural description, where "well-formed" is defined to mean "not containing any appearances of the symbol --". Where it marks the edges of sentences that have not yet been broken down structurally into the basic units, i.e., from the sentence as a whole, a surface structure that, written out, is a "structure" in the surface structure of the sentence, and it is the mechanism that Chomsky proposed to mark as ungrammatical any derivation based on such manipulations as [I saw the boy the dog sat the cat]. Such a structure would "mark" a sentence [that is, I saw the boy the dog sat the cat]; hence a "not technically be a surface structure, and the point here—any underlying structure would not be a deep structure as technically defined."

In this way, certain deep structures were defined derivationally and in a fashion involving cross-level relationships (an effect which we would have to call a derivational constraint, rather than purely in terms of local condition or a given formal representation.

With the passage of time, and a generation of linguists, the notion of derivation changed, in many linguists' perceptions, from being an account of the fundamental problem of levels in linguistics to being the essence itself of linguistic analysis. Now, with the advent of generative semantics and the development of theories of meaning or semantic and structural analysis, that were heavily based on a dynamic model of linguistic analysis, in terms of which one representation is successively changed into another in a sequence that is its entirety is the account of the extraction in question.

In generative syntax, this view has been challenged successfully by approaches such as GSG and, to a somewhat lesser degree, LFG but the derivational view remains powerful and virtually unchallenged in phonology (see, for example, Brenneis and Halle 1988). We must explore why.

There is a close relationship—and frequently a conflation—between the notion of level and that of derivation. The Halle-Chomsky position that challenged the adequacy of the classical phonemic level left a word which the phonological derivation made straight to fill.

Traditional structural phonology allowed three "semantic interpretations": a phonetic representation (PR), a phonemic representation (PM), and a morphophonemic representation (MP). Any expression in a language could be associated with representations on these three levels, and an adequate grammar provided an account of the relationship between...
units on a given level (PT, PM, MP) and those on the other levels. Traditionally, the levels were hierarchized, with rules of allomorphy relating the phonetic and the phonemic levels, and automatic and monomorphic rules of alternation relating the morphophonemic and the phonemic levels. If we indicate allomorphy rules, then, as (PM, PT) rules—rules relating the PM and PT levels, though with no particular significance attributed to the order of these terms—and rules of phonemic alternation as (MP, PM) rules, then there is, quite evidently, an inherent ordering to rules of these two "components," as (2) illustrates.

(2) Morphophonemic Representation
Phonemic Representation
Phonetic Representation
MP
PM
PT

Thus the dismissal of a phonemic level, as argued by Halle and Chomsky, seemed to leave the picture in (2) in a severely weakened position, with only two levels of representation, MP and PT, and thus only one set of principles relating these levels, as in (3).

(3) Morphophonemic Representation
Phonemic Representation
Phonetic Representation
MP
PM
PT

But the minimalist organization offered by (3) is simply inadequate for the treatment of the phonology of any human language, as Chomsky and Halle were prepared to show. In fact, after all, it is necessary to have rule interaction, at the very least of the sort that (2) allows. Let us consider briefly a simple example from a recent analysis of Chukchi by Odden (1985), an example of a sort that could be multiplied ad libitum, to illustrate the kind of interaction that (3) does not permit, but which is very common in language.

Odden demonstrates that there is a complex set of interactions among the phonological rules of Chukchi, including two rules that are in a straightforward feeding order. The first is a rule (4) that changes an r to a before any coronal consonant (including p, which always acts in Chukchi like a coronal consonant). The second is a rule of nasal assimilation (5), which turns stops to nasals before a nasal consonant (we see the effects of this process in the first column, where a schwa has been lost, permitting the nasal assimilation to affect s). Both rules are independently motivated, as the examples provided below, from Odden, illustrate.

(4) Strenthening

<table>
<thead>
<tr>
<th>phonetic</th>
<th>phonemic</th>
</tr>
</thead>
<tbody>
<tr>
<td>pokur-ak</td>
<td>'to arrive'</td>
</tr>
<tr>
<td>zala-ka</td>
<td>'pulled'</td>
</tr>
<tr>
<td>upper</td>
<td>'to haul'</td>
</tr>
<tr>
<td>eger</td>
<td>'star'</td>
</tr>
<tr>
<td>brin-ak</td>
<td>'to buy'</td>
</tr>
<tr>
<td>koy-yo</td>
<td>'puchane'</td>
</tr>
<tr>
<td>informally stated: r ~ r' + x ~ [consonant]</td>
<td></td>
</tr>
</tbody>
</table>

(5) Nasal Assimilation in Chukchi

<table>
<thead>
<tr>
<th>phonetic</th>
<th>phonemic</th>
</tr>
</thead>
<tbody>
<tr>
<td>ye-nama-n</td>
<td>'be ground'</td>
</tr>
<tr>
<td>pese-k</td>
<td>'ground'</td>
</tr>
<tr>
<td>ram-sa</td>
<td>'behinds of hides'</td>
</tr>
<tr>
<td>rapan</td>
<td>'behind of hide'</td>
</tr>
<tr>
<td>ya-nasun</td>
<td>'having news'</td>
</tr>
<tr>
<td>poyl-y</td>
<td>'news'</td>
</tr>
<tr>
<td>ya-namos</td>
<td>'be killed'</td>
</tr>
<tr>
<td>lamp-sak</td>
<td>'kill'</td>
</tr>
<tr>
<td>e-pey-s-ak</td>
<td>'make firewood'</td>
</tr>
<tr>
<td>plya-at</td>
<td>'footwear'</td>
</tr>
<tr>
<td>informally stated: [s] ~ [sonorant]</td>
<td>[sonorant] ~ [nonsonorant]</td>
</tr>
</tbody>
</table>

An example like Jan-sie he bought it (cf. kur-ak 'to buy', in (4)) illustrates that an r before an n becomes a, and then the n becomes a nasal, n; in short, rule (4) creates a structure to which (5) may then apply. What do we say about rule interaction if there are only two levels as phonologica theory? The fact that can be said is that the rules do not directly enter the levels, if there is to be rule interaction of this sort. Ordering, quite necessary in linguistic accounts as we have seen, became then (on this generative account) the function of relations across levels, but a property inherent to another type of object, a rule. That function was to create something that was not a representation on any particular linguistic level. And so was in the real event in the synchronic theory: a representation that was of no particular linguistic level. Such representations are the intermediate forms of generative grammar; they will not be retained here, in the model developed below.

Inherent in the notion of $r-x$ is the idea that there are specific generalizations that can, and perhaps must, be drawn concerning the representation of the expression on that level. For as we saw earlier, a level is a way of representing, or describing, or analyzing, among others, many (perhaps a
boundaries number of levels are thus conceivable, and the correct ones will be justified on the basis of the generalizations that can be drawn, at least, at these few levels.

Synthesizing, then, we may say that the logic of linguistic levels consists in the breaking down of the analysis into a number of distinct, autonomous representations, each with its own elements and its own generalizations (in tactics, or with surface conditions). To this we will add, presently, the notion that each level also contains complexity measures, in such a fashion that certain representations as a given level may be said to be more complex than others on the same level. We will then be in a position to propose that all the functions of derivations be replaced by two types of relations: rules that relate distinct levels, e.g., (MP, PM) rules, and rules that serve to decrease the complexity of representation on a single linguistic level. A grammar that consists of no other form of rule will be called a harmonic grammar.

The notion of derivation in phonology has its origin both in historical linguistics and in the philosophy of mathematics. Historical linguistics, where done as well as Indo-Europeanists know to do it, is based on the possibility of producing a consistent set of pre-language word forms and a set of sound changes consistently ordered in time for each daughter language. Each stage of such a derivation is understood to represent a real time-slice, each stage a (quasi-)phonemic representation of the word in the course of its evolution over time. And for this precise reason, each stage of the derivation has a particular meaning or significance of its own; it directly answers a question, so to speak, such as "What was the phonological shape of the form in A.D. 1200?" It is this characteristic that each stage of the diachronic derivation has in common with a level of linguistic analysis (and, we may add, in common with intermediate stages of a traditional, synchronic phonological derivation); it has specific properties in itself that allow it to be the answer to some question about the expression in question.

The second source of the notion of derivation is generative phonology drawn on the study of the formal properties of proofs in logic and mathematics. The end of the nineteenth century saw the development of the law of reduction of the Kauan notion of synthetic a priori truths—that is, of substantive propositions that could be known to be true without being based on any experience. The very sense of mathematics, the rationalists' image of perfect truth, was placed in jeopardy, and logicians and philosophers asked in what sense traditional Euclidean geometry could be true if non-traditional geometries could be shown to have an inherent truth. It became imperative to study the very nature of mathematical thought—not its syntax as an act of imagination but the logical nature of its context, which is to say, of the production of certain and unsubstantiable conclusions from acceptable assumptions. This project had been begun, of course, by Euclid, but it now was to be extended in a more radical and open-ended fashion. Mathematical proofs themselves would be studied, and become the object of a mathematical analytic: the field would become its own subject of study, for the first time.

A proof has these three properties: first, semantically, each stage contains (or, we might say, preserves) whatever truth inheres in the premise from which that stage has been correctly inferred or deduced; and second, an explicit syntax of inference is defined whereby a stage of the proof can be established—written in, we might say—just in case a specific rule R of inference can operate on one or more of the preceding lines in the proof and produce from them the line in question. For example, given a proposition of the form A or B, we can establish a procedure by which it is syntactically permitted to have the following proposition as; the matter is trivial, but then no set most small, local effects in the syntax of language or anything else.

These various notions were synthesized by L. L. Po in 1943 as the guise of Post production systems, which have the explicit properties just described. Production systems have remained influential in AI work (e.g., J. R. Anderson 1983) and have also, and earlier, served as an intellectual source of the notion of phonological derivation.

Neither the historical derivation nor the Post production system seem, upon present reflection, likely to provide a sound basis for a practical and psychologically valid theory of phonology. Each has evolved with particular ends in mind: historical derivations, to map a long-standing problem of regular neogrammar sound change, and production systems, to account for a formal theory of systems that permits a sentence in which truth must be preserved over the course of particular derivations—while anything, we hardly need add, in phonology. I have emphasized the crucial historicity of the current view of derivations for one simple reason: it is far too easy for us, in our present position, to think of derivations in phonology as arising somehow, inductively and logically, jointly out of the data and the task. It is, rather, in large measure a historical, that is to say, that what the theorems are has most influenced our current views on what is (of which the first six figures in Chomsky's has figured as in this particular view; others are equally so general to the task.
What was lost in the generative emphasis on derivations, and thus equal emphasis upon rules, was the perception of the importance of levels, i.e., that language could be viewed as a whole system of interacting levels of representation. Within recent years, the conception of language as an interaction of abstract and concrete elements, has indeed returned, but too frequently with the emphasis on the rules that function in each part, and the breaking up of the derivation into successive autonomous pieces.

Our central proposal is this: within the phonological part of a grammar, just as within the other components of the grammar, the acquisition of a language consists of the abstraction of a large number of well-formedness conditions—patterns, crudely speaking—on a small number of levels—three, a matter to which we shall return in detail. The proper definition of these language-particular and language-universal patterns will require all of the sophisticated phonological equipment at our disposal; with regard to matters of phonological representation, at least, this harmonic phonology will need all the mechanisms that recent phonological theory has offered.

One more important point may be made now. An objection of considerable generality has been raised to any theory with only three levels of representation by Postal (1968). (His point was made more specifically in the context of criticising Lamb (1966) and others.) Postal emphasizes that in many languages, it is easy to establish derivations—in a traditional generative sense—'with at least four stages, where there is a feeding order among the rules H, first R, feeds R, where

Underlying [Rep₁] [Rep₂] [Rep₃] [Rep₄]
Representation [Rule 1] [Rule 2] [Rule 3]

Thus, if one views a derivation as a path—nearer a plausibly metaphorical sense—from the underlying representation to the surface representation (syntactic phonemic to syntactic phonetic), the path will have at least four links in it; hence there levels will not suffice.

Postal’s argument, we suggest, is supported more by his visual metaphor than by logic. A linguistic level, the reader will recall, is a way of describing a linguistic expression; it thus possesses a vocabulary: a set of features, an inventory of permitted segments (feature combinations), of permitted associations, and so forth. No requirement exists to the effect that the rules or principles of the grammar that establish well-formed representations on any given level must not interact in a feeding fashion. To the contrary: from the suggestions of cognitive psychology, we would expect instead that each linguistic level functions so as to make maximal sense out of the linguistic expression, i.e., to make it maximally conform to its definition of well-formedness.

We may establish, then, for each level L, a set of intralevel (L.L) rules whose function is purely to allow the representation to achieve maximal satisfaction of L’s well-formedness conditions; we may say that the representation at level L relaxes to a maximally well-formed state via the L.L rules, which function as transitions. A level consists of a vocabulary of items, a set of statements regulating how they may be put together, and—it is this third point that distinguishes our proposal from traditional theory—a measure of well-formedness.

From the point of view of traditional generative phonology, the rule applications within a level may be viewed as utilizing the (L.L) rules, but with the following caveats: (1) the (L.L) rules are not (that is, cannot be) ordered; (2) more generally, we will not expect that rules apply sequentially, though for the moment we may make the simplifying assumption that they do; (3) most importantly, intralevel (L.L) rules apply only if their effect is to shift the representation toward a better-formed state; putting the matter differently, the intralevel rules define what the allowable paths are that the representation may move through (or reach) in its quest for a maximally relaxed (well-formed) state.

Even this model is unsatisfactorily derivationally oriented, and in some recent work, which we shall discuss in section 2.3, we have succeeded, explicitly and computationally, in eliminating that aspect. For present purposes, we will be satisfied with the conditions above. The difficulty in avoiding even these remnants of derivationalism lies in the question of how to compose two or more distinct phonological rules, without a more radical revision of the notion of a phonological representation, it is difficult, and perhaps impossible, to treat the effects of several rules without some remnants of derivationalism, which is to say, a linear sequence of distinct, identifiable representations.

We have seen that each level L consists of (1) a vocabulary permitting a linguistic description, (2) a set of relations expressing relative well-formedness, and (3) a set of intralevel (L.L) rules that select the paths that a representation may pass through to find maximal well-formedness. A representation of a given expression on a level L is, in general, not a single representation, but rather a part of representations, i.e., a set of intralevel (L.L) rules Lₜ that defines the (L.L) rules. A grammar, however, consists of a set of perhaps as many as ten or fifteen separate levels, and generally there are specific principles limiting and defining representations on different levels, we will return to these in crucial-level rules (Goldsmith 1990) refers to them as intralevel rules; Sadock 1991.
refers to these as interface principles. We propose, in particular, that these are three levels of phonological interest. (It will not carry out the exercise of illustrating the inadequacy of a system with only two levels, in light of the remarks just above concerning feeding within a single level, the example of feeding in Chukchi does not serve to establish the need for more than two levels, but blending and counterfeature relationships, also common in natural languages, do establish this, and the proposal that there are three levels in, we may assume, the very minimal assumption that could even be considered.) These levels are:

* M-level, a morphophonemic level, the level at which morphemes are phonologically specified;
* W-level, the level at which expressions are structured into well-formed syllables and well-formed words, but with a minimum of redundant phonological information; and
* P-level, a level of broad phonetic description that is the interface with the peripheral articulatory and acoustic devices.

Of these, our greatest interest, as phonologists, will be with the W-level, which is where the bulk of the significant well-formedness conditions, or tactual, are stated. Thence will be six types of phonological rules, to principle: three intrarule types (M,M) rules, (W,W) rules, and (P,P) rules, and three cross-rule types: (M,W) rules, (W,P) rules, and (M,P) rules (we should emphasize again that the order of the symbols is irrelevant; (M,W) rules could as well be called (W,M) rules, for example). This could be depicted as in (6). The traditional hierarchical perception of phonology amounts to a denial of the existence of (M,P) rules, in effect converting (6) to (7), whose form is hierarchical. For the present, we have no need for (M,P) rules, and may assume that such rules do not exist.

We add to the three proposed definitions of levels above, then, the assumption:

* Cross-rule levels may or may not (i.e., need not) be harmonic.

We shall refer to the model in (7) (which is the model we shall explore largely throughout this paper), with its three levels and its assumptions regarding harmonic and arbitrary rule application, as the MW/P model.

2.4. Examples

Let us illustrate these notions with two simple and familiar examples, both well known to practicing phonologists. Their function here is to illustrate the general framework with respect to well-known analyses. The first is the case of Landl, an Australian language discussed by Halle (1973) and several others since, including Kenstowicz and Rosemercy (1979) and also, notably, Ilh (1986) and Williamson (1988), who arrived independently at some similar conclusions, under a different set of theoretical assumptions, the second is the classic case of vowel harmony in Yorubian Yoruba.

2.4.1. Landl

In Landl, whose consonantal inventory is listed in (8), several processes can be seen to be at work in the creation of the uninflected noun, shown in the first column in (9). While the uninflected noun can end in only two ways—with a vowel or with a single apical consonant—a glance at the inflected forms of each noun (here, nounform and future) demonstrates that at some level, the stem of the noun may select one of two more complex sequences of consonants. The level at which this morphophonemic contingency is expressed is, as we have noted, the M-level. Let us consider the M-level forms of the stems, group by group. The forms given here we take to be P-level representations.

![Diagram](image-url)
(8) Consonantal Inventory (Laduli)

<table>
<thead>
<tr>
<th>Laryngeal</th>
<th>Labial</th>
<th>Palatal</th>
<th>Velar</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>j</td>
<td>a</td>
<td>m</td>
</tr>
<tr>
<td>g</td>
<td>n</td>
<td>i</td>
<td>n</td>
</tr>
<tr>
<td>w</td>
<td>y</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(9) Unstressed

<table>
<thead>
<tr>
<th>Nonlaryngeal</th>
<th>Future</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>kej</td>
<td>kej-ε</td>
<td>kej-ε</td>
</tr>
<tr>
<td>miy</td>
<td>miy-ε</td>
<td>miy-ε</td>
</tr>
<tr>
<td>yap</td>
<td>yap-ε</td>
<td>yap-ε</td>
</tr>
<tr>
<td>yaraman</td>
<td>yaraman-ε</td>
<td>yaraman-ε</td>
</tr>
<tr>
<td>jep</td>
<td>jep-ε</td>
<td>jep-ε</td>
</tr>
<tr>
<td>nele</td>
<td>nele-ε</td>
<td>nele-ε</td>
</tr>
<tr>
<td>wanka</td>
<td>wanka-ε</td>
<td>wanka-ε</td>
</tr>
<tr>
<td>koppa</td>
<td>koppa-ε</td>
<td>koppa-ε</td>
</tr>
<tr>
<td>goppa</td>
<td>goppa-ε</td>
<td>goppa-ε</td>
</tr>
<tr>
<td>kupa</td>
<td>kupa-ε</td>
<td>kupa-ε</td>
</tr>
<tr>
<td>gawu</td>
<td>gawu-ε</td>
<td>gawu-ε</td>
</tr>
<tr>
<td>kejje</td>
<td>kejje-ε</td>
<td>kejje-ε</td>
</tr>
<tr>
<td>gije</td>
<td>gije-ε</td>
<td>gije-ε</td>
</tr>
<tr>
<td>pope</td>
<td>pope-ε</td>
<td>pope-ε</td>
</tr>
<tr>
<td>čeŋe</td>
<td>čeŋe-ε</td>
<td>čeŋe-ε</td>
</tr>
<tr>
<td>wije</td>
<td>wije-ε</td>
<td>wije-ε</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Laryngeal</th>
<th>Labial</th>
<th>Palatal</th>
<th>Velar</th>
</tr>
</thead>
<tbody>
<tr>
<td>j</td>
<td>m</td>
<td>i</td>
<td>m</td>
</tr>
<tr>
<td>w</td>
<td>y</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In (9b) the L-level representation is an aster in the gloss-uninflected form: the M-level representation of the inflection suffix is equally uninflected -m, a full form, or for the future. In (9c) the W-level is used to represent the uninflected forms, and the Q-level is used for the unstressed forms. The high vowel that is found final at M-level is intransitive at W-level, and we assume it at W-level as well. This is in the case of a word such as koppa - koppa-ą - koppa-ą - koppa-ą - koppa-ą. In (9d) the M-level is used to represent the M-level representation of the words that are also vowel-final in the inflected forms before the inflection suffixes but instead of corresponding to a non-high vowel on the W-level, by (10), as in the forms, the word-final vowel is entirely missing on the W-level. This cross-linguistic pattern holds for all forms where there are two or more vowels on the M-level; i.e., where there are two or more vowels on the W-level despite the effect of cross-level rule (10).

(10) M-level: m

<table>
<thead>
<tr>
<th>L</th>
<th>W</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>m</td>
<td>m</td>
<td>m</td>
</tr>
</tbody>
</table>

(On this type of rule formulation, see Lakoff, Chapter 4 of this volume.)

This correspondence is expressed by a cross-level rule—here, (M,W) rule—given in (11).

(11) M/W | m/W | m/W

In (11) we see that the M-level representations of these words are also vowel-final: the word-final vowel is realized in the inflected forms before the inflection suffixes, but instead of corresponding to a non-high vowel on the W-level, by (10), as in the forms, the word-final vowel is entirely missing on the W-level. This cross-linguistic pattern holds for all forms where there are two or more vowels on the M-level. In (12) the W-level is used to represent the M-level representation of the words that are also vowel-final in the inflected forms before the inflection suffixes, but instead of corresponding to a non-high vowel on the W-level, by (10), as in the forms, the word-final vowel is entirely missing on the W-level. This cross-linguistic pattern holds for all forms where there are two or more vowels on the M-level; i.e., where there are two or more vowels on the W-level despite the effect of cross-level rule (10).

(12) M-level: m/W

<table>
<thead>
<tr>
<th>L</th>
<th>W</th>
</tr>
</thead>
<tbody>
<tr>
<td>m</td>
<td>m</td>
</tr>
</tbody>
</table>

Harmonic Phonology 35
The non-indication of a correspondence in (12) means correspondence is identity. Any form in which (12) is relevant will also be relevant to (11), but in the case of such conflict, the more specific rule—here, (12)—has priority.

In (9b), we see examples of stems ending in a consonant at M-level which is not present at P-level; thus the form for 'shack' (paran, paranpos, paraosh) clearly ends in an 'q' which appears at P-level in the two inflected forms, but does not appear at P-level in the uninflected form. Similarly for gulk, whose P-level representation for the uninflected form is gulk.

All languages place heavy importance on syllabification at the W-level, and the effects we see here are the result of such conditions in Lard (noted also by Wilkins 1988). The heavier restrictions, here as elsewhere, are placed on what is allowed in a syllable code position. Lardite permits no more than one single apical consonant in a word-final code. Thus, all of the uninflected form nouns in the first column represent well-formed words at W-level as well as P-level, but such forms as 'narkys', for example, do not:

(13) i. apical

Thus the transition within W-level of (i) to (ii) represents a shift toward greater word formation, in that each segment must, in a well-formed W-level sense, be licensed by an appropriate licensor (such as a syllable or an appendix).

The offending, or orphaned, p in (i) is deleted in the transition from (i) to (ii) by rule (14). This intrarelative (WR) rule applies harmonically—i.e. only when its effect is to create a better-formed representation, or, alternatively put (as we sound above), the rule defines a path, or transition, that a representation may embark upon, but the representation will take that direction if and only if it constitutes a relaxation, or improvement, for that representation. If we interpreted (14) as a production rule—in a traditional generative rule—its conditions for application would be entirely contained within its structural description, and it would apply if, and only if, its structural description were met.

(14) C* → of [__] (WR)

In a harmonic system, a rule such as (14) indicates a (language-particular) permissible change, but such a change is efficient only if this provides an increase in well-formedness, such as more properly meeting the condition that at W-level, all segments are properly licensed.

Harmonic Phonology 37

The rule in (14) is thus purely harmonic in function, unlike both (11) and (12). As further examples in (9c) illustrate (e.g., myrnaparap), a high vowel may appear word-finally at W-level; the shift in (11) is thus not motivated by a w; i.e., a condition on permissible W-level representations (12) appears, similarly, not to be motivated by an effort to satisfy a phoneticism on the W-level, though Wilkins (1988) has suggested in effect that it is; we leave it mute open here. These differences illustrate an important point.

(15) All intrarelative rules are harmonic in function; cross-level rules may or may not be harmonic in function.

The forms in (9b) illustrate the logical reading of the (W,W) rule (14) by the (MW,W) rule (12). In the case of a form like M-level shakb nazir 'husband', the (MW,W) rule is responsible for the loss of the final vowel, and the (W,W) rule allows the resultant shakp nazir to get rid of its final, offending p:

(16) M-level y u k a r p

W-level y u k a r p

y u k a r

The evolution of the W-level representation—here, the loss of the final p—is entirely parallel to the construction of syllable and metrical structure. Rather than drawing the most fundamental distinction between, e.g., structure-building and structure-changing operations, as in lexical phonology, we here draw the most important distinction between operations moving in a direction that holds generally across a level (roughly, that is transparent), on the one hand, and those that relate representations on distinct levels, on the other.

The derivation of a form such as miyraparap (from miyraparape) illustrates the double application of the (W,W) rule (14).

2.4.2. Yakuts

Let us turn now to the second example, that of Yakutsian Yakuts. Our discussion here is based on the now well-known interpretation of S. Newman's work by S. Y. Kudrya 1967 and by C. Kerschhoff 1969; most recent work in this area includes Archangelski 1985, and one may also see Hockett 1973 for an overview of earlier work and Dell 1972 for a good introduction. Lukoff, in this volume, presents a discussion of much of the same material within a similar framework as well. Yakuts, like the Ladin example to a burst examine,
present novel familiar material which understandably serves as a litmus test for the plausibility of an approach to phonology of analysis.

These facts have been cited in the past to illustrate, first, the necessity of rule ordering—one may argue that five linearly ordered rules must be pos-
tioned here—and, second, that the inventory of phonological phenomena (vowel, *vowels*) shifts during the course of the generative derivation—in our terms, the inventory of vowels is different at the W level and at the P-
level, and, in a slightly more abstract sense, different also at the M-level.

Looking ahead, we find in Yukut a 4 x 2 vowel system at the M- and W-levels (17a), and a 2 x 2 vowel system at the P-level (17b).

(17) Vowel inventory (Yukut)

<table>
<thead>
<tr>
<th>a</th>
<th>i</th>
<th>u</th>
<th>short</th>
</tr>
</thead>
<tbody>
<tr>
<td>e</td>
<td>o</td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>i</td>
<td>u</td>
<td>long</td>
<td>(I), (U)</td>
</tr>
</tbody>
</table>

a: e: u

The materials we are primarily concerned with are given in (18), which we may take as representing the stable, or final, state of the P-level repre-
sentation.

(18) Future Ancst Geosynclinal Dolostonic Glass

<table>
<thead>
<tr>
<th>a</th>
<th>CVC</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>xil</td>
</tr>
<tr>
<td>u</td>
<td>xil</td>
</tr>
</tbody>
</table>

xilna xilal *xiln* (a situation)

b: CVCV

<table>
<thead>
<tr>
<th>a</th>
<th>xil</th>
</tr>
</thead>
<tbody>
<tr>
<td>u</td>
<td>xilal</td>
</tr>
</tbody>
</table>

*recognition*

gapal gopal *gop*

*a take care (of a child)*

maxat maxan maxa: *obtain*

b: CVCV

<table>
<thead>
<tr>
<th>a</th>
<th>mokal</th>
</tr>
</thead>
<tbody>
<tr>
<td>u</td>
<td>mokal</td>
</tr>
</tbody>
</table>

*mokal* *swallow*

songat sogn: *songal* *unpack*

dontat dontan donat *remove*

tannat tannan tannal *go*

C: CVCV

<table>
<thead>
<tr>
<th>a</th>
<th>hweorg</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>hweorg</td>
</tr>
</tbody>
</table>

*somewhere* *walk*

sudokul sudokuk sudokal *remove*

jopek ul jopekl jopek *get up*

yesawal yesawat yesawal *follow*

d: CVCV

<table>
<thead>
<tr>
<th>a</th>
<th>lei</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>lei</td>
</tr>
</tbody>
</table>

*expose to wind*

cyounat cyoun: cyounal *unain*
role of vowel harmony can be straightforwardly established which copies rounding to a vowel to the rule of the same height, as in (21). The formulation in (21) assumes (perhaps wrongly, see Archangeli 1985) that consecutive identical specifications of vowel height have been collapsed (i.e., an OCP effect; see Goldsmith 1990, 309ff.), and that the vowel specifications for the four-vowel system of (18a) are as in (22); that is, the feature Low is equipollent and the feature Round is privative. This assumption is necessary if we are to have a simple and direct formulation of Round spreading, as in (21); the notion of two vowels being at the same height can only be expressed if both height specifications play a role in the phonology (i.e., if the feature is equipollent), and spreading of rounding can most simply be stated if the target of the spreading—the unrounded vowel that "becomes" round—is not specified for the feature Round.

(21) Rounding Harmony [\[n \text{ Low} \]]

(22) \begin{align*}
\text{Low} & : u \quad a \quad o \\
\text{Round} & : \quad +
\end{align*}

The effect of Rounding Harmony, thus, is to create a situation in which the span of the feature Round matches up, as far as possible, with the feature Low (the significance of this alignment of two spreading domains was first proposed in Steriade 1981; it has been developed as well in Archangeli 1985, where this point was first made regarding Yamalnaiiuyu, under a different set of assumptions regarding features, and also in Cole 1987, where the Yoruba case, among others, is discussed further). This kind of matching is more general than the generative-style notion in (21) suggests. If we are to succeed in separating the goal state of this level, on the one hand, from the path permitted to reach it, on the other, we will need an effective and efficient system for indicating appropriate goal states. One such common goal is the matching in terms of a sequential association over two distinct charts. We may express this "harmony" between the charts F and G as: F \to G. If we specify that the chart Round consisting of the tiers Low (shaded), round, and the chart Height (shaded, low) are harmonic in the sense that Round \Rightarrow Height, then the rule given in (21) may be simplified to (23), a harmonic rule that applies only if it increases the well-formedness of the representation. In this particular case, we take harmony between charts, when specified in the grammar, to contribute to overall well-formedness, a point emphasized in the works by Steriade, Archangeli, and Cole referenced above.

(23) Rounding Harmony (revised)

\[ [n \text{ Round}] \]

This well-known type of analysis of Yawelmani Yakut vowel harmony is based, as we have said, on the premise that some goal e between e and u corresponds to any long e on the P-level as a long e on the M-level, corresponding to some long a's on the P-level as long u's on the M-level (other P-level a's correspond to M-level a's). We have no basis as yet for determining whether the correspondence of e: with e (and u: with a:) occurs between the M- and W-level, or between the W- and P-level. Put another way, does the W-level have the 4 \times 2 vowel system of (24a) or the 5 \times 2 vowel system of (24b)?

(24) Long Vowel Lowering

\begin{align*}
\text{a:} & : \ U \text{ M-level} \quad \text{b:} & : \ U \text{ M-level} \\
\text{e:} & : \ W \text{ W-level} \quad \text{u:} & : \ W \text{ W-level} \\
\text{e:} & : \ P \text{ P-level} \quad \text{u:} & : \ P \text{ P-level} \\
\text{[M, W] rule:} \quad \text{[W, P] rule:} & \\
\text{M:} & \Rightarrow W \quad \text{P:} & \Rightarrow W \\
\text{[low] \Rightarrow [low]}
\end{align*}

We cannot be certain yet. The three rules we have so far are insufficient to motivate any properties for a level other than the M-level and the P-level; in fact, our three rules so far—Closed Syllable Shortening (20), Rounding Harmony (23), and Long Vowel Lowering (24a or b)—can be perfectly well expressed in a much simpler model with two levels (M and P); we might call them (here), in which the three rules directly control the two-levels, no issues of footing, blending, or ordering are necessary; the rules operate independently—what has been called "minimally.

When we consider the rule of Epenthesis, however, we find that the situation comes more sharply into focus. The fact that it does not always work with a pattern with roots whose M-level form is CVVCVC. When a consonant-initial
Harmonic Phonology

The shift from the free-vowel system of (22) to that of (29), as in (30), is thus part of the surface relation associating the W-level and the P-level. Again, it is worth emphasizing that this relationship should not be thought of as a familiar derivational one, in which the W-level representation is somehow trying to "get out" and only do so by running through the maze of phonological rules and constraints. Each level, rather, sets its own autonomous conditions, and the relationship in (24) is thus parallel, e.g., to the syntax/phonology interface, or a D-level/PS-level relationship. In the interface sketched in (30), there is no conversion in the
character of both features: Low is equivalent on one level, private on the other, and the same holds for Round.

(30) M-level, W-level

<table>
<thead>
<tr>
<th>[round]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[low]</td>
</tr>
<tr>
<td>[low]</td>
</tr>
<tr>
<td>[low]</td>
</tr>
</tbody>
</table>

P-level

<table>
<thead>
<tr>
<th>[round]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[low]</td>
</tr>
<tr>
<td>[low]</td>
</tr>
</tbody>
</table>

The feature interfaces utilized in (30) are both interesting and straightforward, and deserve to be shifted into feature type noted just above.

(i) Round is equivalent at M, W-level, but equivalent at P-level;

(ii) [round] at P-level interfaces with [round] at W-level, while [round] at P-level interfaces with B at W-level (as, of course, does B at P-level). These are the default correspondences that are imposed by the change in feature type.

(iii) Similarly, Low is equivalent at W-level and private at P-level; again, [low] at W-level interfaces with [low] at P-level, while [low] at W-level interfaces with B at P-level. In general, feature shifts of this sort across levels follow the natural pattern given in (32).

(32) If M, W-level vowels are subject to the level-particular restriction that the (equivalent) feature Low must be specified for each vowel, then no vowel will directly interface with the P-level, as desired. The lowering rule of (i) may be reinterpreted as a (W,P) rule, as in (32), or in a quite different fashion, as the result of a P-level tactic specifying that all long vowels (i.e., vowels associated with two positions, nuclear and coda) are associated with [low], i.e., cannot be the vowels (or a, see (33)).

(33) P-equivalent

<table>
<thead>
<tr>
<th>P: private</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1,</td>
</tr>
<tr>
<td>+F</td>
</tr>
<tr>
<td>-F</td>
</tr>
</tbody>
</table>

Harmonic Phonology 45

From a familiar generative point of view, the structural differences between the M, W, and P levels may seem odd, off putting, and embarrassing for the model. To repeat, this intuition, we believe, is solely the result of the thoroughgoing power of the derivation metaphor—the production line metaphor, if you will, applied to a linguistic object. There is no object moving or being translated from one of the three levels to another; what is "happening" during the linguistic analysis is that there (simultaneous) types of analysis are being compared, contrasted, and measured for fit from this point of view, further structural differences between levels are quite welcome.

Archangeli (1985) discusses another important matter that we interpret here as such a structural difference between levels. She suggests, in essence, that the vowel patterns of stems as in (18c-d), with an "echo" or copied, vowel in the second syllable, is the result of a morphological spreading. We may illustrate this as in (34), for the forms honorou and tapiou.

(34) a. h y w v y v V - V b. t y p y v v V - V

| [low]   |
| [low]   |
| [low]   |
| [low]   |
| [low]   |

There is universal consensus among analysts of Yukon that such echo stems are to be analyzed with repetitions or copies of the same vowel in the two vowel positions of the stems (thus: *aten, *joke, *hype, *jiuk, *nuk, etc.), and autosegmental theory requires (on one view, to which I subscribe; it strongly suggests, on another view) that the copying is done by
autonomous spreading illustrated in (34), however, the rule of Lowering only involves a single syllable in Yawelmani Yokuta (as other dialects, see Archangeli, 1983). At level a, this corresponds to P-level Lowering, not, after all, hower it (or hower it). This suggests that the representation to which Lowering (24) applies is "syntactically local," as in (35), in a way in which the representations in (34) are not.

(35) a. 

![Image](image.png)

b. [x x] [x x] [x x]

\[
\begin{array}{ll}
V & V \ V \ V \\
V & V & V & V \\
\text{low} & \text{low} & \text{low} & \text{low}
\end{array}
\]

In the local representations of (35), each syllable has copies (to speak) of the relevant autosegmenta (i.e., stems). While we have no specific additional grounds, we may speculate that this symmetrical splitting of features is in some part due to the restructuring of the feature types (equivalent, primitive) across W- and P-levels.

We have explored in this section the shift in perspective that arises when we attempt to decompose a phonological analysis into intrasegmental and cross-level components, and to reinterpret the familiar and traditional derivational analyses in terms that emphasize the tactics specific to autonomous levels of the phonological component.

Through this discussion, an insatiable remnant of the derivational conception has remained in our intrasegmental analysis, in the sense that even the free and unordered application of rules within a level reconstructs a major part of the notion of derivation, though with the differences noted above. In the next section, we turn to a new conception, that of a dynamic model, which overcomes this defect.

2.5 Dynamic Computational Models

In this section, we turn to a new direction, and offer an explicit approach that addresses directly the aspect of the MWP harmonic model mentioned just above, the residue of the linear derivation within the intrasegmental rule applications.

We generally view the typical interaction of two (or more) rules as the composition of these rules, as if the output of one were to provide the input specifications for the other. We do this, in general, for two reasons: first, we often discover feeding relationships between two processes which, from an analytical point of view, appear to be conceptually distinct, as in the case of the two rules in Chukche mentioned above, cf. (14), (5). Second, while a coherent definition of simultaneous application is possible under certain conditions, it is often difficult to ascertain whether those conditions will be met by a given set of rules. If the rules operate on sharply distinct subsets of the representation, they may apply simultaneously—but it may be difficult to prove that two rules are in principle independent in that respect, and, worse yet, if we have a set of n rules which are independent in that sense, it is in general not easy to determine whether adding some specific n + 1st rule will be possible without significant interaction arising.

Composition of rules—guaranteed by our traditional notions of rule ordering—is a simple way out of this technical morass of potential rule interaction and conflict.

In the previous section, I suggested that intrasegmental rules applied freely, with free reapplication. Such a suggestion naturally raises questions such as whether one could be certain that no ambiguities would arise; whether, for example, a case might not arise in which a rule of openness fed a rule of deletion, which in turn fed the openness, which in turn fed the deletion, ad infinitum. One might equally attempt to demonstrate that free rule application of intrasegmental rules would not lead to unwanted and incorrect ambiguities of output, depending on which of two rules R₁ and R₂ was "chosen" at a given point to apply when the structural description of each was satisfied.

In this section I will sketch an approach I have been exploring more recently that appears to be able to resolve many of the potential difficulties just raised by shifting the nature of the rules in question, that is, by shrank- ing the computational distance between the representations, on the one hand, and the rules, on the other. Gary Larson and I refer to models of this approach as dynamic computational models, and within the broader grammatical picture sketched in the previous sections, we take them to be models of the individual levels of the MWP model, along with the intrasegmental rules of the level in question. A single dynamic computational model, that is, represents simultaneously the representations, the tactics, and the intrasegmental rules of a particular level.

There are three essential properties to these models. The first involves more a style of presentation than of rigid substance, but it is a precondition for the second and third properties. Rather than viewing representations as concatenations of formal objects (segments, autosegments, or constituents composed of these) with various features assigned to them in which one feature specification may be algebraically replaced by another to form a new representation, we design a fixed device which can take on various
states, each of these states is a phonological representation. A representation
is viewed, then, as a state of the device, and in some cases—the representation—can be thought of as a point in a space of high-dimensional, n-.
ue space. To be sure, more traditional modes of phonological thought
can be described in such terms not uncomfortably as well, but we will
make use of this mathematical perspective directly.

Second, the units each take on an activation value, typically in the
range from 0 to 1, but not restricted to that range. Specific phonological
properties are not, then, inherently binary or quantized in any way, though
categorical properties may arise out of "winner take all" arrangements
among certain subgroups of units, or other similar constructs involving
threshold values.

The third, and final, property of the dynamic computational models is that
units communicate their activation values to their neighbors in ways that
are in part universal and in part language specific. As they do so, the
state of the model as a whole shifts, and thus its character as a specific
phonological representation; the communication of activation from one
unit to another causes the device to shift from representational state S_j
to representational state S_k. Under conditions that can be made precise, this
evolution of the system due to activation passing converges to a fixed
point, or equilibrium, and this is the output of the system. Each such sys-
tem represents a phonological level of the sort discussed in the previous
section.

In a dynamic computational model, there is no sharp contrast between
the rules and the representations as such. Roughly speaking, rules (in the
traditional sense) correspond to the network of connections that pass ac-
tivation from one unit to another, and to the specific arithmetic coefficients
assigned to those connections. This correspondence is rough, to be sure.
The effects of familiar phonological rules cannot always, or in general, be
simply reformulated in such quantitative terms. But perhaps the single
most important effect of translating all dynamic effects into a quantitative
system is that we can deal with all interactions of conflicting demands as a
matter of arithmetic; conflicts are resolved, so to speak, by simply let-
ting the stronger influence dominate. The final, or equilibrium, state of
the device is the result of all the component forces simultaneously
recognized.

In the remaining pages, I shall briefly sketch how a system of this sort
operates in practical terms. More details are available in Goldsmith (in press a,b), Goldsmith and Larson 1990, Larson 1990, and we refer
the reader as well to Goldsmith (to appear a) Goldsmith and Larson (in
preparation).

We focus here on two central concerns of phonological theory, accent
systems and syllabification systems, for two reasons: first, these are sys-
tems that are highly productive and regular in the, in, that the vast majority
of languages of the world, and second, these are systems in which rule in-
teractions and long derivatives have been the rule rather than the exception
in discussions of the

One basic model is the same for the treatment of both accent and syllab-
ification systems. We consider a linear sequence of units $a_i$, each of which
has an activation value $x_i$ which may be positive or negative. These units
are ordered roughly from the bottom units of the metrical grid in the case of
the accentual system, and in the skeletal tier in the case of the syllabifica-
tion device. Each unit has an inherent activation, which is the sum of two
terms: the internal activation (corresponding to syllable weight in the case
of the accentual system, and to sonority in the case of the syllabification
operations on the skeletal tier) and positional activation—monotonic as-
signed on the basis of position in the linear string.

Furthermore, each unit $a_i$ is connected to its left and right hand neigh-
bors in such a way that the more activated $a_i$, the more it inhibits its
neighbors, $a_{i-1}$ and $a_{i+1}$. Conversely, a unit with negative activation ex-
cites its left and right hand neighbors; we refer to this relationship, there-
fore, as one of polarity inhibition.

Let us make these relationships a bit more precise. We may imagine,
for exposition purposes, that the activation of all of the units is updated
at the same time, i.e., synchronously. The activation $x_i$ of a given unit $a_i$ at
time $t$, or, $x_i^t$, and its value is determined by equation (36), where $\alpha$ and $\beta$ are coefficients between 0 and −1 (though more generally $\alpha$ and $\beta$
can be positive as well) specifying the degree to which an element affects
its left and its right-hand neighbor, respectively:

$$x_i^{t+1} = (1 - \alpha \beta) x_i^t + \alpha (x_{i-1}^t + x_{i+1}^t)$$

In general, the absolute value of the product of $\alpha$ and $\beta$ must be less than
about 0.5 for the system to be stable and well-defined. Typical values
for $\alpha$ and $\beta$ are $\alpha = -0.35$ and $\beta = -0.1$, i.e., $1 - \alpha \beta$, though in
the case of accentual systems there is reason to expect the value of $\beta$ to refer
to the derived activation of the units.

These assumptions have already built something much like the rule of
Perfect Grid right into the architecture of the metrical system. If a single
unit $a_i$ favors, for whatever reason, has a positive inherent activation, then
its left and right-hand neighbors will have negative activation; this is likely
true, will be positively activated; and the activation of unit $a_{i-1}$ will spread
out dynamically through the network. For example, if the fifth unit in a

0.2
system with six units has an inherent activation of 1.0, with $\alpha$ and $\beta$ set at 0.8; thus the system will reach an equilibrium with the following values:

$$
\begin{align*}
\text{unit 1} & \quad \text{unit 2} & \quad \text{unit 3} & \quad \text{unit 4} & \quad \text{unit 5} & \quad \text{unit 6} \\
0.08 & \quad -0.16 & \quad 0.31 & \quad -0.50 & \quad 1.11 & \quad -0.11
\end{align*}
$$

In general, stress or accent is assigned to those units whose activation is a local maximum, that is, whose activation is greater than that of their neighbors on either side. In addition, a threshold value ($\beta$) may be established below which even a local maximum is simply ignored.

As we noted above, this quantitative approach allows for a simple notion of dynamic interaction of phonological influences. Consider, for example, the interaction of positional activation and polar inhibition. In the more familiar terms of metrical phonology, we may ask how the system deals with a typical case such as the following: Suppose that a language stresses the penultimate syllable, with alternating stress being caused by factors from that position, and that in addition, the first syllable is stressed. How do these rules interact, and how is the potential conflict between the two rules resolved, particularly in the case of a word with an odd number of syllables? Will the prepenultimate, the fourth syllable from the end, always be stressed? Will it be stressed if that would create a clash with the first syllable? Will the first syllable "lose" its stress because of comparison with stressed syllables to its right?

Stress on the penultimate, as well as stress on the first syllable, is inherent positional activation. We assume that all inherent positional activation is quantitatively the same: 1.0. Perfect Grid accounts, within this model, for any settings of $\alpha$ and $\beta$ below 1.0, and a low setting for $\beta$, the threshold value. The resultant values can be calculated straightforwardly. If we assume that the values of ($\alpha$, $\beta$) are -0.5 and -0.1 respectively, then the values we get for three, four, and five syllable words are as in (38); this is based on a calculation in which an inherent activation of 1.0 is assigned to the first and to the penultimate syllables. In a system in which the local maximum is phonetically realized as stressed, the syllables that are marked with an accent will be stressed.

$$
\begin{align*}
\text{3 syllables} & \quad 4 \quad \text{syllables} \\
0.5 & \quad 1.0 & \quad -0.1 & \quad 1.35 & \quad -0.7 & \quad 1.3 & \quad -0.11
\end{align*}
$$

While these are the values of ($\alpha$, $\beta$) that appear to be correct in most of the systems we have looked at, the type of systems available is larger, and other equilibrium values are accessible with other settings of $\alpha$, $\beta$ (the setting of these values for the accent system is, however, a constant for a given language). This is illustrated in figure 2.1 for the case of a three-syllable word. In region A, the first syllable $x_1$ is a local maximum, and $x_2$ is greater than $x_2$ in region B, $x_i$ is greater than $x_i$ in region C, and $x_j$ is greater than $x_j$. We see there, in a system in which, by varying the explicit parameters, we can shift from one accent system to another, is the sense that within the distinct regimes, distinct accent patterns are found for certain words or word types, but the general and crosslinguistic properties of accent systems, such as stress-sha avoidance and alternating stress, are the result of the inherent architecture of the system.
that even cyclic action, by treating the simultaneous poss. of a language with, for example, the effect of cyclic def by assigning sc. that is true in Goldsmith-Lazard, is the main thing that a movement of a deriv. that is correct for the results of the tests that we have, we do indicate how this direction comes into play for both, and in terms, act systemic in the system described above, however, when we look for peaks, or聆 events, it is not peaks of inherent sonority that we seek, but rather of what we call derived sonority. Assuming values of a and b in the neighborhood of 0.5 and 0.4 respectively, the high level of sonority (ut sonority) of the elements \( x_i \) decreases the activation of its left-hand neighbor, \( x_{i-1} \), while itself, greatly reduced in activation, affects the activity of its right-hand neighbor, \( x_{i+1} \), very little; thus the derived activation of the sequence \( x_{i-1} x_i x_{i+1} \) declines from \( x_{i-1} x_i \) to \( x_i \) and then increases from \( x_i \) to \( x_{i+1} \), and in general, the observed syllabic nuclei will coincide with the derived sonority.

Lazarow has explored in Lazard 1990 and in the paper the ways in which the saliency predictions of the Minimum Saliency Distance Principle (Searle 1982, Harris 1982, and much recent work) can be derived from checking no more than the locations of peaks and troughs in the curve of derived sonority, in this area, while additional results follow as well, as Lazard (1990) discusses in connection with the syllabification pattern of Spanish. Without going into great detail here, some of the general features of the model can be mentioned. Just as a peak of derived sonority corresponds to the nucleus of a syllable, so the local minimum (or trough) of sonority marks the boundary between syllables; in general, syllables are stretched from one trough of derived sonority up to, but not including, the next trough of sonority.

Let us calculate when a sequence of segments \( x_i, x_{i+1} \), may constitute an onset-nucleus sequence. That will be possible if the derived sonority (ds) of the arguments satisfies the conditions in (39):

\[
\delta(x_i) < \delta(x_{i+1}) < \delta(x_{i+1})
\]

We would like to know what conditions this sets on the inherent sonority of the arguments \( x_i, x_{i+1} \), which arguments will be permitted to appear in such a sequence? Given any arbitrary values of \( a \) and \( b \), the relations...
between inherent and derived sonority is a bit complex, but we can begin to
get a feel for the effect if we consider a first approximation for the
situation where $\beta = 0$, and where we calculate only the first (but most significant)
inception in the step toward equifinal. Under those conditions, the first
inequality in (39) reduces to (40), where $x$ is the inherent sonority of
segment $\xi$.

\begin{equation}
(40) \quad (x - a_x) < (x - a_y), \quad \text{that is,} \quad x_y - x > a(x - x_y)
\end{equation}

Thus, what this model suggests is that a minimum sonority difference is
required between successive elements in the onset as such, but rather
that a strict upper limit is set on the difference between the sonority of
the second and third element, a limit determined by the difference in sonority
between the first and second. (Larson 1990 discusses this effect in greater
detail.)

Let us consider the treatment of a sequence of two consonants between
vowels a bit further. While a sequence such as /sl/... will be syllabified
unnecessarily as /...l/... in virtually all languages, the sequence
/aba... is, as is well known, syllabified in some languages as /...ab... and
in others as /...a.../. The present model calculates the syllable division
partly on the basis of the determination of the local minimum of de-
manded sonority, and it is not difficult to observe that as we adjust the $\alpha$
parameter from 0 up toward 1.0, we observe the syllable division per-
formed in these two different ways. With $\alpha = 0$, $b$ is a sonority mini-
mum; as $\alpha$ increases, $\alpha$ sonority decreases, until it falls below $b$'s
at a high enough value of $\alpha$. Precisely where this point is depends on
the assignment of inherent sonority values to the individual segments, a matter
as to which we now turn.

In work currently in progress, Larson and I have explored the properties
of a system that abstracts the sonority values in the following way. We
present the device with a list of words from a given language, specified in
a familiar (binary) feature system, and we indicate which segments are in
fact, on the surface, the syllable nuclei, i.e., the peaks of derived sonority.

The device then proceeds to calculate sonority in the following fashion. It
assumes that the total sonority of a segment is the sum of the contribu-
tions to sonority of each of its individual features, and for simplicity's
sake, we assume that one value of each feature contributes no sonority.

The other value of the feature will contribute to the sonority of the seg-
ment—but how much? For every feature $\xi$ there is a corresponding
amount $\xi$, and the task of determining the sonority hierarchy for a given
language is then equivalent to determining the values of $\xi$ for all the fea-
tures $\xi$. We may refer to the set of coefficients $\xi$ as the sonority vector $S$.

Our results so far suggest that this is not a difficult task for even a
simple device. Our device assumes a fixed value of $\alpha$ and $\beta (0.5-0.2)$, and it begins (or, rather, we begin) with purely random values of the set of coefficients $\xi$. Then it calculates a sonority curve, with its maxima and minima, on the basis of the three totally random sonority coefficients, and compares its gustars of where the sonority maxima are with the data as we
have presented it—since we inform the device which segments are the
(surface) sonority peaks. When in gusters is correct, the coefficients are not
changed. When the device gustars that a segment is a peak but it is not, then it decreases by 10 percent the coefficients $\xi$ that contributed to its
sonority, and conversely, when it fails to identify a true sonority peak, it
increases by 10 percent the coefficients $\xi$ of the features that contribute to
the sonority of that segment. In short, it acts like a perceptual.

For our first dorer words, the device does no better than chance in the
contrast performance. Its performance quickly improves, however. For example,
when we use the two hundred most common English words as a training set,
the device correctly assigns syllable peaks to 62 percent of the words after one trial, and to 97.5 percent after five trials.

This kind of performance, rudimentary though it is, suggests how an
expect learning procedure of this sort can allow a formal device to arrive
at the correct sonority vector for a given language. We discuss this issue at
greater length in Goldsmith and Larson (in prep.)

Finally, as we noted above, the development of dynamic computational
models for both accent and syllabification systems offers the prospect of
developing a deeper account of the relationships between the two systems.

There are two fundamental relationships between the two: (1) it is peaks,
and only peaks, of derived sonority that correspond to elements of the ac-
ccentual system (we may say that an accentual unit $a$ is associated with a
skelatal unit $\xi$, or that they are associates, just in case $\xi$ is a local maxi-
mum), (2) in what are called quantity-sensitive systems, the activation, or
total amount of (derived) sonority, of a skelatal unit contributes to the acti-
vation of its associate on the accentual tier (Pitts 1982).

In work in progress, Larson and I are exploring the nature of quantity-
sensitive accentual systems (discussed in detail, for example, in Hale and
Vergnaud 1987) by establishing local dynamic links between the local
maxima on the skelatal tier and their associates on the accentual tier. In
doing so, we have integrated into the heart of the accentual model the
two central properties of accentual systems of the world: their propensity
to promote alternate syllables to prominence (introduced in the world in
the local connections of polar inibition, as in (36)), and their propensity
to assign greater accentual prominence to those syllables incorporating
Harmonic Phonology

Notes

1. This chapter has been written in collaboration with colleagues over the past several years. It develops ideas that are discussed in detail in Gold-smith (1989,1990, in press a), and has been influenced by many linguists and non-linguists that I could name. With regard to the notion of h-score, I have been especially influenced by Charles Hockett and Norman Chomsky, as well as by J.R. Firth and Kenneth Pike; the reader will note the echoes of semantical, phonemic, and interpretive thinking in some of the discussion herein. I have had a number of dis- cursos with George Lakoff since 1987 on the notion of rule ordering and the challenges to intermediate representation. On the nature of harmonic rule application, the work by Singh and Zomeister that I cite below deserves particular mention, as do its other ways, lecturers by John McCarthy at the 1987 Linguistics Institute at Stanford University. With regard to lexical and harmonic phonology, various works on lexical phonology and phrase level phonology have been influential in each of the work in these traditions seems to me to be quite wrongheaded; the present work aims to provide more attractive alternatives. On the importance of the notion of harmonic application in a larger cognitive perspective, I have been greatly encouraged by current work in constructionism, as discussed, for example, in Keesbergen and McCollough et al., 1986, and I have found especially helpful the work on the work of Sokolinsky (1986 and elsewhere), whose work in harmony theory was at the origin of my choice of name for the present work. I have also been influenced by the work of Sidney Lamb and Charles Hockett; I should note. I am greatly indebted to my colleagues and students at the University of Chicago, especially to Jerry Ladd, whose work on aesthetical theory, among other things, behind the present discussion, and in discussions with the audience in our phonology seminars, including Anna Bock, Diane Brouwer, Gary Larson, Karen Penston, and Candace Wilterdink. Geoffrey Hark, Younghee He, and Jesse Puhin, have made very helpful comments on earlier drafts of this paper. This material was presented at the Institute of Phonetics and the University of California at Berkeley, and I am grateful to the linguists there for helpful comments as well.

2. This material is based upon work supported by the National Science Foundation under Grant No. BM7-3000678.


4. Perhaps the phrase is underdetermined: There are no morphological principles given in the phrase that would allow one to guess when a theory’s phonology have acquired too far; too far what can be translated into a psychologically accurate model?

5. To my surprise (and dismay), I have found these preliminary observa- tions are occasionally guided with a certain hostility among my otherw ise em- otional and sensible colleagues, I would explicitly avoid giving the impression either that I am rejecting all of contemporary phonological thought (that is the hardest thing for me) or that I am suggesting that my proposal here is without precedent in the more literature. On the contrary, it is rather obvi-
ous, as I have tried to suggest here and in Goldsmith (1980, chap. 8), that the present proposal is a natural development of proposals that have been made over the past twenty years, and more, in some cases. I would emphasize that the central points of the present discussion is to encourage discussion, within the present context, of the relationship between derivations and levels. A version of the proposals in the current literature bear directly on the model discussed here. In certain respects, the "repair strategy" explication of Sacks (1984) and, following him, Pandolfo (1988), based explicitly on the propositional structure in generative syntax (1974), are the closest, but to other respects their proposals are quite distinct (or perhaps "orthogonal" would be a better term) to the present discussion. Tap (1988) explores similar ideas. As I have noted, the work by Sydney Lamb, mentioned exceedingly rarely in the present literature despite its influence, is to be read. Some work of Alan Prince, of John McCarthy, and of John Lohb is highly relevant, though the list of work that is highly relevant would be hard to complete once it was started.

3. Throughout this section I use the vague pronoun "we" to refer to a common group of contemporary linguists. The reader may take the word to be used either exclusively or exclusively as the reader wishes; my intention is that it be inclusive.

4. A recent and thoughtful discussion of these notions is found in Lakoff 1988.

5. For purposes of conciseness, I take Hockett's position in Hockett 1961 to epitomize the theoretical position that I refer to here as American structurism. Chomsky 1975/1975 offers a fine expression of the notion of level.

The development of a system of phonetic representation does not conclude the process of linguistic analysis. We also want to discover the morphemes, words, and phrases of the language, and to discern principles of sentence structure that could hardly be stated directly in terms of phonemes. Instead of giving a direct definition of these further notions within linguistic theory, we can continue to construct systems of representation for sentence structure, calling these systems "linguistic levels." A sentence often can be represented as a sequence of phonemes, but it can also be represented as a sequence of morphemes, words, and phrases. Thus each sentence taken will have associated with it a whole set of representations, each representation being its "spelling" in terms of elements of one linguistic level. (Pp. 98-99)

Perhaps better known is the following passage, on the opening page of Chomsky 1975:

The central notion in linguistic theory is that of "linguistic level." A linguistic level, such as phonemes, morphology, phrase structure, is essentially a set of descriptive devices that are made available for the construction of grammars; it constitutes a certain method for representing unification. (P. 11)

6. It goes beyond the scope of this paper to discuss the matter in detail, but the point may well be made that a major component in the lack of communication between the generative semantics and the interpretive semantics is the late 1960s derived from the set, by the generative semantics, of the internalized notion of derivations without a serious commitment to the view of syntax that is the linkeffect method of note. Global rules (Labov 1963, 1967) seem like a formal abbreviation from the point of view of derivations, although, (not from the point of view of a theory linking sentence semantics with syntax). Syntactic and semantic levels of analysis in terms of a fashion as possible---"the aspect" is the notion of an ambiguous level of deep structure, distinct from semantic representation, was not included in the theory. This issue is discussed further in Goldsmith and Block 1981.

7. A reviewer poses the question at his point as to whether we are interpreting the relations between levels as being achieved by rules working simultaneously at each level, i.e., (MP)(PM) and (PM)(PT). The answer is yes, of course, though the structuralist would say that the word "simultaneity" is out of place. A relation is not something that happens; it is a state, we might say, not an event, and all cross-level relations are timeless states. If 2 + 0 equal 2, and 0 = 3 X 4, are we committed to saying that the two equations are simultaneous when we write 2 + 0 = 2 X 4 in a word, no, to say they are one simultaneously would be misleading at best, a category mistake at worse.

8. Anđla Komlósi informs me that he has recently argued that such a position---developing a semantics, in effect, for phonology---is interesting, in a number of cases, I have not had the opportunity to study this work (Komal 1993).

9. Chomsky 1975/1975 provides a discussion of the introduction of representation that are mentioned only by general simplicity of rule interactions (pp. 114-16 especially).

10. The present proposal leaves open a number of questions. If intonational rules, for example, are considered, in the claim not implicit that no cases will arise in which the order of application of two intonational rules must be specified in a language-particular fashion. We will not, in fact, address that question, because we are trying to establish a rather different way of discussing phonological rules, alluded to in the final section of this chapter. The purpose of the approach is that what we take to be phonological rules here be formalized as something approaching a field of force in a large-dimensional space; these fields can be automatically measured, and the sum is what determines the effect that the phonological rule undergoes.

11. This formal foundations must be interpreted as being the equivalent of a Structural Description and a Structural Change; the double assumption of a feature in the 3D, and the base is what there must be. More is to be said about this distinction.


13. The grammars that rule applications will be kept well from a formal point of view have been elaborated and the in cases found in the literature in the pronouns and appears as they might be expected from rules, persistent processes that regulate the repetition by scoring structure or ensuring that it is well-formed in the first place such being the Well-Formedness Condition of Goldsmith
60 John Goldsmith

1979 [1976]: The danger to be found in the incorrect use of such devices is this: Suppose we have an ordered set of rules, R', and a set of fewer than n total pressure repair processes, P. After one of the rules (say, R_j) applies, all of the pressure re-

pair processes might apply. But do they apply immediately or do they apply sequentially? If sequentially, how do we know which one has first chance—will universal principles determine that, or will the repair processes be ordered in a language-particular fashion? In either case, we will be left with two sets of or-

dered procedures, one for applying inside the derivation established by the other set. To my knowledge, none of these bizarre consequences have been addressed in the literature, but the problems for current theory are very real. Hill [1986], in a major development in this line of research, is led to establish two pressure processes, Syllabification and Synt Right, a number of her examples, such as the discussion of Lautst, illustrate the requirement that after a (real) phonological rule has applied, Synt Right must not apply until after Syllabification has ap-

plied, but after that, the two processes apply freely, in alternating fashion, until the structural condition for neither of them is met. My point is not that this sort of process interaction is impossible, nor even that it is inextricable in any sort of way, but rather that familiar rule-oriented phonological theory has—on the basis of current and valid generalizations—backed itself into a corner where radical re-

thinking of rule organization is simply a necessity.

16. Of course, many of these art phonological representations are not states of the device need be defined as in its physiological representarions.


18. This sentence leaves open whether end units on either the left or right-

hand end of the word are local maxima when their activation is greater than that of their (single) neighbors. Either choice on this—saying that they do count as local maxima, or that they do not—would be reasonable, and this factor appears to us at this point to be a language-particular decision. When an element must be strictly greater than elements on either side, then no peripheral element will be a local maximum no matter what its activation value, and we derive what is known in the literature as peripheral extrametricality (extraproximity).

The maximum in the text is also incomplete in that it may be a matter of a number of cases to take stress to the as-

signed to these elements whose activation is above a language-specific threshold, regardless of whether the unit is a local maximum or not. The reader familiar with the subject may note that this distinction is directly related to whether or not the language allows stress clash.

19. This is not a general assumption across the theory, in fact, the option of varying the relative strength of inherent activation is an important parameter within the approach.

20. Except for Scass Caeli; see Bondus 1991; here syllabaf are defined as streches from (but not including) one tough up to and including the next minimum.