2

Harmonic Phonology

John Goldsmith

To appear in ***The Last Phonological Rule*: *Reflections on Constraints and Derivations***, edited by John Goldsmith. University of Chicago Press, 1993.

2.1. Introduction

This chapter is a discussion of several basic issues which I have been exploring recently.[1] All of them involve very simple questions, so simple that they will no doubt frequently run the risk of striking the reader as having been fully and satisfactorily treated in the past. I am convinced that this is not so, and can do no more than invite the reader to reconsider some of these questions with me. Among the issues I am concerned with are the matter of extrinsic rule ordering; the appropriateness in a derivation of intermediate stages which are not at a specifiable linguistic level; and, most importantly, the notion of the derivation as a sequential set of steps, as a part of a production system. In section 2.2. I describe more recent work done in collaboration with Gary Larson, developing a dynamic model, an explicit computation model that offers a more radical interpretation of intralevel phonological processes.

Much of the material discussed here arose out of a critical analysis of lexical phonology,[2] which depends heavily on what appear to be thoroughgoing uses of an implausible metaphor[3] involving space and time: "First add an affix, then send that material through a set of rules which modifies the resultant form; then go to the next level, add another affix, and finally string all the words together, only after which do we reach a point where the postlexical rules get a chance to apply." The hope is implicit in such an account that the ungainly metaphors are present only for expository reasons; but as I attempted to extract the essence from the packaging, for my own purely pedagogical purposes, I slowly, and reluctantly, came to the conclusion that the operation left little behind. In short, and at the risk of oversimplifying, the essence of léxical phonology emerged as an implausible metaphor. But such a conclusion demands swift and positive action: lexical phonology has the good sense to make us confront important questions, and we must not lose sight of that. More than that, there are some important insights at the core of lexical phonology that seem to easily get lost—such as the thorough-going identity of morpheme-structure conditions and lexical phonological rules.

The phonological conclusion that the present paper aims toward is this: that all phonological rules which apply at a particular level have the explicit function of moving a representation as far as possible toward meeting 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 (*pace* suggestions of structure preservation); and, finally, stratificationalism: 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 propaedeutic to a proper and full treatment.[4]

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 three are potentially problematic notions, and the boundaries are on occasion difficult to define. But this tripartite division is nonetheless very useful, and worth the effort we expend on establishing it.

Of the three notions, that of representation is the most familiar at present. Most of the work in phonological theory 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 autosegmental tiers and association lines, of metrical trees and grids, of feature geometries, of dependency relations among feature specifications, of "particulate" 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 American phonological thought had been that phonological representation was largely unproblematic and consisted, in particular, of sequences of 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 theoretical phonologists during this period. Much of this has changed now, to be sure, but the change has been a recent 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 this has been in large part a matter of focus of attention: as 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 concerns 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[5] are all comfortable with the notion of levels in linguistic analysis; but I invite the reader to consider some basic questions once again.[6] Our first goal is to clarify the notion of level, and when we do so, we find, first, that it is best (and quite well) explicated in pregenerative writings; second, that generative grammar was originally conceived—no surprise!—as an answer to questions formulated within this clear and traditional understanding of the notion of level; and third, that the notion of a derivation has passed from being, at its origin, a possible approach to the problem of linguistic levels, to a view of linguistic analysis which stands in the way of a clear understanding of the notion of levels.[7]

Like all important ideas, that of linguistic levels is very simple: a linguistic level is a way of looking at—of describing—a linguistic expression. We may look at an expression from a syntactic perspective, and posit a syntactic level of representation; or from a morphological perspective, and posit 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 on the syntactic level (always, or typically) correspond to one or more 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 ordering is not inherent 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 morphemes, and so as to be less misleading, it is often best not to spell a morpheme as, say, *dog* or [dɔg], but rather as μ_{122}, for example, to emphasize that it has no internal segmental structure at that level. The analysis of an utterance such as *the dog barked* into morphophonemic elements, which may well include boundaries (if our theory countenances them) separating the morphemes, as in (1c), does not constitute a representation on the morphological level—it is an analysis on a phonological or morphophonological level, as evidenced precisely by the kind of units (here, phonological units) that constitute the representation.

(1)	a. The dog barked.	(English orthography)
	b. [det: definite] [μ_{345}] [Verb$_{533}$] [PAST]	(morphological level)
	c. ðə+dɔg+bark+d	(morphophonemic level)

Levels, 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; Hockett (1961), for example, argued at length that the morphological level and the phonological (or, specifically, phonemic) level are autonomous and cannot be viewed as having places on a single linear hierarchy of levels. Pike's work generally addressed this question, in the context of a theory involving three hierarchies of distinct levels; see Pike 1972 and references therein. Of this literature, little seems to remain in linguists' collective memory beyond the specific phonological issues formulated by Chomsky (1964), and yet the issues remain of significance.

Generative phonology—and generative grammar, more generally—proposed a specific account of the relation among traditional linguistic levels, an account that centers around Chomsky's 1975 [1955] conception of a linguistic derivation. Certain pairs of levels were to be related to each other by means of derivations (specifically, the systematic phonemic and the systematic phonetic levels, as well as deep structure and surface structure); in other cases, the level itself was defined in terms of a derivation. In a sense, this last move turns the logic of analysis into levels on its head, in a way that deserves our close attention. The most striking example of this is found in *Aspects of the Theory of Syntax* (Chomsky 1965, 138), where the notion of deep structure is defined not simply as the result of lexical insertion into the output of the phrase structure rules, the picture we tend to remember; a further condition is set there on being a deep structure. A structure which is created by the phrase structure rules and to which the

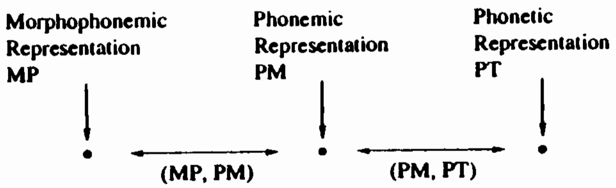
rules of lexical insertion successfully assign lexical elements is called a "generalized phrase-marker." A generalized phrase-marker is a deep structure only if, in addition, it is part of a derivation that includes a well-formed surface structure, where "well-formed" is defined to mean "not containing any appearances of the symbol #"—where "#" marks the edges of sentences that have not yet been integrated derivationally into the larger, matrix clause. Thus a surface structure that contains a "#" is no surface structure at all, and it is this mechanism that Chomsky proposed to mark as ungrammatical any derivation based on such monstrosities as [*I saw the boy* [*the dog bit the cat*]]. Such a structure would "surface" with a telltale "#" (that is, as *I saw the boy #the dog bit the cat#*); hence it would not technically be a surface structure, and—the point here—its underlying structure would not be a deep structure as technically defined. In this way, the notion of deep structure was defined derivationally and in a fashion involving cross-level relationships (an effect which we would later learn to call a derivational constraint), rather than purely in terms of local conditions on a given formal representation.

With the passage of time, and of a generation of linguists, the notion of derivation changed, in many linguists' perception, from being an account of the fundamental problem of levels in linguistics to being the essence itself of a linguistic analysis. No precise moment dates this transition, but the late 1960s saw the development of modes of speaking about syntax and about phonology that were heavily committed to a dynamic model of linguistic analysis, in terms of which one representation is successively changed into another in a sequence that in its entirety is the account of the expression in question.

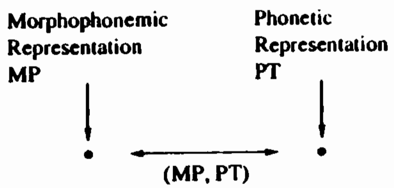
In generative syntax, this view has been challenged successively by approaches such as GPSG and, to a somewhat lesser degree, LFG.[1] But the derivational view remains powerful and virtually without challenge in phonology (see, for example, Bromberger and Halle 1989). We must explore why.

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

Traditional structuralist phonology allowed three quasi-phonological representations: a phonetic representation (PT), a phonemic representation (PM), and a morphophonemic representation (MP). Any expression in a language could be associated with a representation 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 allophony relating the phonemic and the phonetic levels, and automatic and nonautomatic rules of alternation relating the morphophonemic and the phonemic levels. If we indicate allophony 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.[5]

(2)


Thus the dismissal of a phonemic level, as argued by Halle and Chomsky, seemed to leave the picture in (2) in a seriously 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)


But the minimalistic 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 case after case, 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 (1988), 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 straightforwardly feeding order. The first is a rule (4) that changes an *r* to a *t* before any coronal consonant (including *y*, 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 a stop). Both rules are independently motivated, as the examples provided below, from Odden, illustrate.

(4) *r*-Strengthening

pəkir-ək	'to arrive'	pəkit-tək	'you(pl.) arrived'
qeper	'glutton'	qapat-čəŋ-ən	'big glutton'
eŋer	'star'	aŋat-ḷəŋ-ən	'star'
kur-ək	'to buy'	kot-yo	'purchase'

Informally stated: r → t / ____ [+coronal]

(5) Nasal Assimilation in Chukchi

Nasalized		Non-nasalized	
γe-nme-ļin	"he ground"	pəne-k	"grind"
rəmn-ət	"flesh sides of hides"	rəpən	"flesh side of hide"
γa-mŋət-ļen	"having news"	pəŋəļ	"news"
γa-nmə-ļen	"he killed"	təm-ək	"kill"
te-pļeŋ-ŋ-ək	"make footware"	pļek-ət	"footware"

Informally stated: $[-\text{sonorant}] \rightarrow \begin{bmatrix}+\text{sonorant}\\+\text{nasal}\end{bmatrix} / \_\_\_\_ [+\text{nasal}]$

An example like *kun-nin* 'he bought it' (cf. *kur-ək* 'to buy', in (4)) illustrates that an *r* before an *n* first becomes a *t*, and then the *t* 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 in phonological theory? The least that can be said is that the rules do not directly relate 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) not the function of relations across levels, but a property inherent to another type of object, a rule, whose function was to create something that was not a representation on any particular linguistic level. And that was a very special innovation in linguistic 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 level 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 said earlier, a level is a way of representing, or describing, or analyzing, an utterance; many (perhaps a boundless number of) levels are thus conceivable, and the correct ones will be justified on the basis of the generalizations that can be stated at just those 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 (or tactics, or well-formedness conditions). To this we will add, presently, the notion that each level also contains complexity measures, in such a fashion that certain representations on 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 sort 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, when done as well as Indo-Europeanists know to do it, is based on the possibility of providing a consistent set of protolanguage 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, a surface (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. 1350?" It is this characteristic that each stage of the diachronic derivation has in common with a level of linguistic analysis (and, we may add, not 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 in generative phonology draws on the study of the formal properties of proofs in logic and mathematics. The end of the nineteenth century saw the overthrow of the Kantian 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 status of mathematics, the rationalist's image of perfect truth, was placed in jeopardy, and logicians and philosophers asked in what sense traditional Euclidean geometry could be true if nontraditional

geometries could be shown to have no inherent flaws. It became imperative to study the very nature of mathematical thought—not its nature as an act of imagination but the logical nature of its content, which is to say, of the production of certain and indubitable 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 analysis: the field would become its own subject of study, for the first time.

A proof has these two properties: first, semantically, each stage contains (or, we might say, preserves) whatever truth inheres in the premises 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_i 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 & b*, we can establish a procedure by which it is syntactically permitted to have the following proposition *b;* the matter is trivial, but then so are most small, local effects in the syntax of language or anything else.

These various notions were synthesized by E. Post (1943) in 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 mature reflection, likely to provide a sound basis for a practical and psychologically valid theory of phonology. Each had evolved with particular ends in mind: historical derivations, to treat a long-standing problem of regular Neogrammarian sound change, and production systems, to account for a formal theory of systems that possess a semantics in which truth must be preserved over the course of particular derivations—unlike anything, we hardly need add, in phonology.[10] I have emphasized the peculiar 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, ineluctably and logically, jointly out of the data and the task. It is, rather, in large measure a historical, or even biographical, matter that the theoreticians who have most influenced our current views on this (of whom the first and foremost is Chomsky) have offered us this particular view; others are equally congenial 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.[11] Within recent years, the conception of language as an interaction of autonomous components, or modules, 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 put—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 general significance 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 criticizing 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: R_1 feeds R_2 feeds R_3 where

Underlying Representation	⟶ [Rule 1]	Rep_2	⟶ [Rule 2]	Rep_3	⟶ [Rule 3]	Rep_4

Thus, if one views a derivation as a path—in some presumably metaphorical sense—from the underlying representation to the surface representation (systematic phonemic to systematic phonetic), the path will have at least four links in it; hence three 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 a mini-derivation, 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 search) 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 2.5, we have succeeded, explicitly and computationally, in eliminating that aspect.[12] 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 which express the paths that a representation may pass through to find maximal well-formedness. A representation of a given expression on a level L is thus, in general, not a single representation, but rather a pair of representations (L_i, L_f) (i.e., initial and final) where L_f is the best-formed (with respect to L's phonotactics) representation accessible to L_i, given 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 refer to these as *cross-level rules* (Goldsmith 1990 refers to them as interlevel rules; Sadock 1991 refers to them as interface principles). We propose, in particular, that there are three levels of phonological interest. (I 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 bleeding and counterfeeding relations, also common in natural languages, do establish this; thus the proposal that there are three levels is, 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.

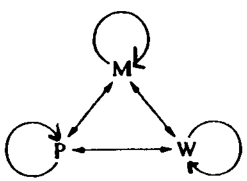
Of these, our greatest interest, as phonologists, will be with the W-level, which is where the bulk of the significant well-formedness conditions, or tactics, are stated.

There will thus be six types of phonological rules, in principle: three intralevel rule types: (M,M) rules, (W,W) rules, and (P,P) rules, and three cross-level 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 conception 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.

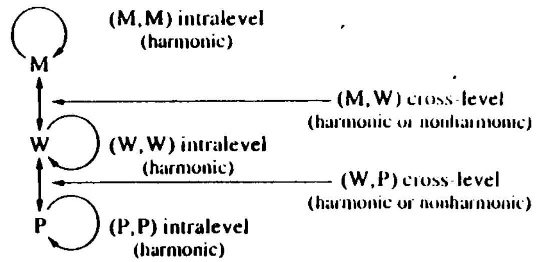
We add to the three proposed definitions of levels above, then, the assumption:

- Cross-level rules may or may not (i.e., need not) be harmonic.

(6)



(7)



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 ***M/W/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 Lardil, an Australian language discussed by Hale (1973) and several writers since, including Kenstowicz and Kisseberth (1979) and also, notably, Itô (1986) and Wilkinson (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 Yawelmani Yokuts.

2.4.1. *Lardil*

In Lardil, 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, nonfuture and future) demonstrates that at some level, the stem of the noun may indeed end in more complex sequences of consonants. The level at which this morphemic constancy 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.

(8) Consonantal Inventory (Lardil)

Labial	Laminal dental	Apico-alveolar	Laminal alveolar	Apico-domal	Velar
p	t̪	t	t'	ṭ	k
m	n̪	n	n'	ṇ	ŋ
		l	l'		
		r		ṛ	
w			y		

(9)

	Uninflected	Nonfuture	Future	Gloss
a.	kentapal	ketapal-in	kentapal-uṛ	'dugong'
	keṭar	keṭar-in	keṭar-uṛ	'river'
	miyaṛ	miyaṛ-in	miyaṛ-uṛ	'spear'
	yaṛput	yaṛput'-in	yaṛpuṭ-uṛ	'snake, bird'
	yaraman	yaraman-in	yaraman-kuṛ	'horse'
	pirŋen	pirŋen-in	pirŋen-kuṛ	'woman'
b.	mela	mela-n	mela-ṛ	'sea'
	wanka	wanka-n	wanka-ṛ	'arm'
	kuŋka	kuŋka-n	kuŋka-ṛ	'groin'
	ŋuka	ŋuku-n	ŋuku-ṛ	'water'
	kaṭa	kaṭu-n	kaṭu-ṛ	'child'
	ŋawa	ŋawu-n	ŋawu-ṛ	'wife'
	kent̪e	kent̪i-n	kent̪i-wuṛ	'wife'
	ŋin̪e	ŋin̪i-n	ŋin̪i-wuṛ	'skin'
	pape	papi-n	papi-wuṛ	'father's mother'
	t'empe	t'empe-n	t'empe-ṛ	'mother's father'
	wit̪e	wit̪e-n	wite-ṛ	'interior'
c.	yalul	yalulun	yaluluṛ	'flame'
	mayar	mayaran	mayaraṛ	'rainbow'
	wiwal	wiwalan	wiwalaṛ	'bush mango'
	karikar	karikarin	karikariwuṛ	'butter-fish'
	yiliyil	yiliyilin	yiliyiliwuṛ	'oyster (sp.)'
d.	ṭurara	ṭuraraŋin	ṭuraraŋkuṛ	'shark'
	ŋalu	ŋalukin	ŋalukuṛ	'story'
	putu	putukan	putukaṛ	'short'
	murkuni	murkuniman	murkunimaṛ	'nullah'
	ŋawuŋa	ŋawuŋawun	ŋawuŋawuṛ	'termite'
	tipiti	tipitipin	tipitipiwuṛ	'rock-cod (sp.)'
	ṭapu	ṭaput'in	ṭaput'iwuṛ	'older brother'

	muŋkumu	muŋkumuŋkun	muŋkumuŋkuṛ	'wooden axe'
	t'umput'u	t'umput'umpun	t'umput'umpuṛ	'dragonfly'
e.	yukar	yikarpan	yukarpaṛ	'husband'
	wulun	wulunkan	wulunkaṛ	'fruit (sp.)'
	wuṭal	wuṭalt'in	wuṭalt'iwuṛ	'meat'
	kantukan	kantukantun	kantukantuṛ	'red'
	karwakar	karwakarwan	karwakarwaṛ	'wattle (sp.)'

In (9a) the M-level representation is as seen in the given uninflected form; the M-level representation of the inflectional suffixes is equally unproblematic: *-in* for the nonfuture, *-uṛ* for the future. In (9b) the P-level uninflected forms given differ from their M-level representation in that a high vowel that is word-final at M-level is nonhigh at P-level (and, we presume, at W-level as well). That is, in the case of a word such as *kent̪e –kent̪in–kent̪iwuṛ* 'wife', there is a high vowel *i* at M-level corresponding to a nonhigh vowel at W-level and P-level:

(10) M-level $i]_{word}$
|
W-level $e]_{word}$
|
P-level $e]_{word}$

(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 [V] $]_{word}$
↕ |
W [−high]

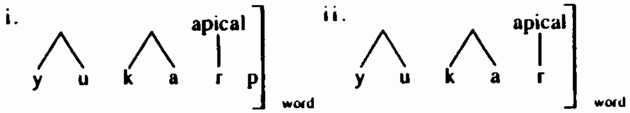
In (9c) we see that the M-level representations of these words are also vowel-final—the final vowel being realized in the inflected forms before the inflectional suffixes—but instead of corresponding to a nonhigh vowel on the W-level, by (10), as with (9b) forms, the word-final vowel is entirely missing on the W-level. This cross-level pattern holds for all forms where there are three 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 (12).

(12) M-level V C V C V$]_{word}$
↕ |
W-level ø

The non-indication of a correspondence in (12) means correspondence is identity. Any form to 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 (9d), we see examples of stems ending in a consonant at M-level which is not present at P-level; thus the form for 'shark' (*ţurara, ţuraraŋin, ţuraraŋkuṛ*) clearly ends in an *ŋ*, which appears at P-level in the two inflected forms, but which does not appear at P-level in the uninflected form. Similarly for *ŋaluk*, whose P-level representation for the uninflected form is *ŋalu*.

All languages place heavy conditions on syllabification at the W-level, and the effects we see here are the result of such conditions in Lardil (noted also by Wilkinson 1988). The heaviest restrictions, here as elsewhere, are placed on what is allowed in a syllable coda position. Lardil permits no more than a single apical consonant in a word-final coda. Thus, all of the uninflected noun forms in the first column represent well-formed words at W-level as well as P-level, but such forms as *yukarp*, for example, do not:

(13)


Thus the transition within W-level of (i) to (ii) represents a shift toward greater well-formedness, in that each segment must, in a well-formed W-level structure, be licensed by an appropriate licenser (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 intralevel (W,W) rule applies harmonically, i.e., only when its effect is to create a better-formed representation; or, alternatively put (as we noted 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—as 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.

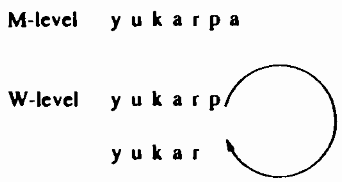
(14) $C \rightarrow \emptyset$ / ____] (W,W)

In a harmonic system, a rule such as (14) indicates a (language-particular) permissible change, but such a change is effected 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.

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

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

The forms in (9d) illustrate the logical feeding of the (W,W) rule (14) by the (M,W) rule (12). In the case of a form like M-level *yukarpa* 'husband', the (M,W) rule is responsible for the loss of the final vowel, and the (W,W) rule allows the resultant *yukarp* to get rid of its final, offending *p*:

(16)


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 *muŋkumu* (from *muŋkumuŋku*) illustrates the double application of the (W,W) rule (14).

2.4.2. *Yokuts*

Let us turn now to the second example, that of Yawelmani Yokuts. Our discussion here is based on the now well-known interpretation of S. Newman's work by S.-Y. Kuroda 1967 and by C. Kisseberth 1969; more recent work in this area includes Archangeli 1985, and one may also see Hockett 1973 for an overview of earlier work and Dell 1973 for a good introduction. Lakoff, in this volume, presents a discussion of much of the same material within a similar framework as well. Yokuts, like the Lardil example to a lesser extent,

c1966 ch2.qk 10/18/92 8:04 PM Page 38

presents now familiar material which understandably serves as a litmus test for initial plausibility of an approach to phonological analysis.

These facts has been taken in the past to illustrate, first, the necessity of rule ordering—one may argue that five linearly ordered rules must be posited here—and, second, that the inventory of phonological elements (here, 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 Yokuts a 4 × 2 vowel system at the M- and W-levels (17a), and a 5 × 2 vowel system at the P-level (17b).

(17) Vowel inventory (Yokuts)

a. i	u	short		b. i		u	short
a	o			e	a	o	
i:	u:	long		(i:)		(u:)	long
a:	o:			e:	a:	o:	

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 representation.

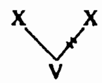
(18)

Future passive	Aorist passive	Gerundive	Dubitative	Gloss
a.	CVC			
xilnit	xilit	xil?as	xilal	'confuse (a situation)'
hudnut	hudut	hud?as	hudal	'recognize'
gopnit	gopit	gop?os	gopol	'take care (of a child)'
maxnit	maxit	max?as	maxal	'obtain'
b. CVVC				
meknit	meekit	mek?as	meekal	'swallow'
sognut	soogut	sog?as	soogal	'unpack'
dosnit	doosit	dos?os	doosol	'recount'
tannit	taanit	tan?as	taanal	'go'
c. CV CVVC				
hiwetnit	hiweetit	hiwet?as	hiweetal	'walk'
sudoknut	sudookut	sudok?as	sudookal	'remove'
?opotnit	?opootit	?opot?os	?opootol	'get up'
yawalnit	yawaalit	yawal?as	yawaalal	'follow'
d. CV CVV				
?ileenit	?ilet	ilee?as	?ilel	'expose to wind'
cuyoonut	cuyot	cuyoo?as	cuyol	'urinate'
hoyoonit	hoyot	hoyoo?os	hoyol	'name'
panaanit	panat	panaa?as	panal	'arrive'
e.	Epenthesis CVCC			
?iliknit	?ilkit	?ilik?as	?ilkal	'sing'
?ugunnit	?ugnut	?ugun?as	?ugnal	'drink'
logiwnit	logwit	logiw?as	logwol	'pulverize'
?ayiynit	?ayyit	?ayiy?as	?ayyal	'pole a boat'

Syllable structure plays an important role in Yokuts at both the W-level and the P-level. It is at the P-level (and, as we shall see, not at the W-level) that the very common CVX syllable structure is found. In such a CVX system, all syllables are of the form CV (with a short vowel), CV: (with a long vowel), or CVC (a syllable closed by a single consonant or glide). At the P-level as well, we find the canonical 5 × 2 vowel system of (17b). There are, however, any number of cases where a long vowel is followed by two consonants, as in the case of a form like *taan* 'go' when followed by the suffix *-nit*, as in (19).

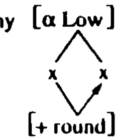
(19)


Such a string cannot be syllabified at P-level into CVX syllables, and a rule of closed syllable shortenings (20) applies harmonically (i.e., as needed) to shorten vowels to achieve proper syllabification at P-level (we return below to whether (20) is best formulated as a (P,P) rule or as a (W,P) rule).

(20) V: ⟶ V or:


What has made Yawelmani Yokuts the testing ground for phonological theories is a rule of vowel harmony whose effect is to spread rounding to the right, within words, among vowels of the same height regardless of length. This generalization, though, is only accurate at a level of representation removed at a certain degree from the P-level representation of (18). Looking at the forms in (18a), built from forms containing only short vowels at M-, W-, and P-level, we see the simple pattern clearly illustrated. A suffix *-it, -nit* will have the corresponding round vowel (*u*) after a stem with *-u-* in it (e.g., *hud*), and nowhere else; a suffix *-al, -?as* will have a round vowel in it after a stem with *-o-* in it (e.g., *gop*). The forms in (18b) can only be understood if we take the M-level representations to be, respectively, *mi:k, su:g, do:s, ta:n;* on the basis of these representations, a rule of vowel harmony can be straightforwardly established which copies rounding to a vowel to the right 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 unround vowel that "becomes" round—is not specified for the feature Round.

(21) Rounding Harmony

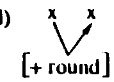


(22)

	i	u	a	o
Low ~~High~~	–	–	+	+
Round		+		+

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 Yamalmani Yokuts, under a different set of assumptions regarding features, and also in Cole 1987, where the Yokuts case, among others, is discussed further). This kind of matching is more general than the generative-style notation 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 span of autosegmental association over two distinct charts. We may express this "harmony" between the charts F and G as: F ≈ G. If we specify that the chart Round consisting of the tiers {Skeleton, round} and the chart Height {Skeleton, low} are harmonic in the sense that Round ≈ Height, then the rule given in (21) may be simplied 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)



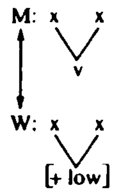
This well-known type of analysis of Yawelmani Yokuts vowel harmony is based, as we have said, on the premise that what corresponds to any long *e:* on the P-level is a long *i:* on the M-level; corresponding to some long *o:*'s on the P-level are long *u:*'s on the M-level (other P-level *o:*'s correspond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

(24) Long Vowel Lowering

a.

i:	u:	M-level
e:	o:	W-level
e:	o:	P-level

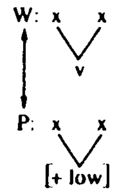
(M,W) rule:



b.

i:	u:	M-level
i:	u:	W-level
e:	o:	P-level

(W,P) rule:



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 only two levels (M and P, we might call them), in which the three rules directly mediate the two levels; no issues of feeding, bleeding, or ordering are necessary; the rules operate independently—what has been called "simultaneously."

When we consider the rule of Epenthesis, however, we find that the situation comes more sharply into focus. The forms in (18e) illustrate a pattern with roots whose M-level form is CVCC. When a consonant-initial suffix follows, a triconsonantal cluster is produced, as in (25). In such cases an epenthetic *i* appears, clearly in order to allow a representation that conforms to a proper syllabification.

(25) M-level ʔil k-nit
ʔilik- nit

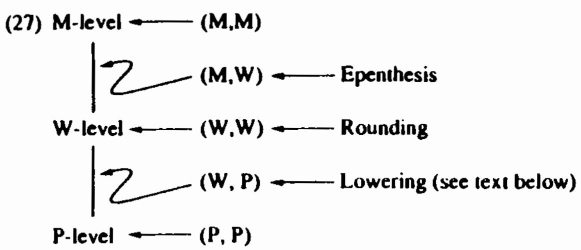
This epenthetic *i*, however, has a clear relationship to vowel harmony: first, such an epenthetic *i* will be made round by Rounding Harmony (23) (cf. (26a)); and second, an occurrence of an epenthetic *i* will serve to block an occurrence of Rounding Harmony (23) between an *o* and a following *a* at M-level (26b).

(26) a. ʔugn-nit	M-level	b. so:nl-ʔas
ʔugin-nit		so:nil-ʔas
ʔugun-nut		*so:nilʔos

In generative derivational terms, Epenthesis both feeds and bleeds Rounding Harmony; there is a clear and extrinsic ordering established between them. From the point of view of harmonic phonology, this effect can be the result only of the placement of the two rules in different groups of the M-/W-/P-model, with Epenthesis "higher" than Rounding in the picture in (27).

Furthermore, Epenthesis cannot be an (M,M) rule, the "highest" grouping available, because epenthesis applies in a harmonic fashion to achieve well-formed syllabification—in the event, to get rid of unsyllabifiable sequences of three consecutive consonants—and syllabification is not a well-formedness condition in Yokuts, nor in any language, at the M-level.

Thus Epenthesis may be an (M,W) rule (or a (W,W), (W,P), or (P,P) rule), and hence Rounding may be neither an (M,M) nor an (M,W) rule; it may only be a (W,W), (W,P), or (P,P) rule, as in (27).

(27)


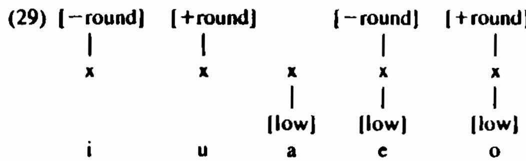
We may now draw a conclusion regarding whether the vowel system at W-level includes the long high vowels (24b) or not (24a). Rounding

Harmony must operate on a level at which the long high vowels are found (otherwise, Rounding would not correctly apply to a form such as *sog-nut*, with M-level *su:g-nit*); hence W-level must have the 4 × 2 vowel system of (24a), and hence (24b) is correct. That is, Lowering is a (W,P) rule, as in (27). We cannot determine whether Shortening (20) is (W,P) or (P,P).

(28) Level	Vowel System	
M-level	4 × 2	su:g-nit
W-level	4 × 2	su:g nit > su:g nut
P-level	5 × 2	so:g nut > sog nut

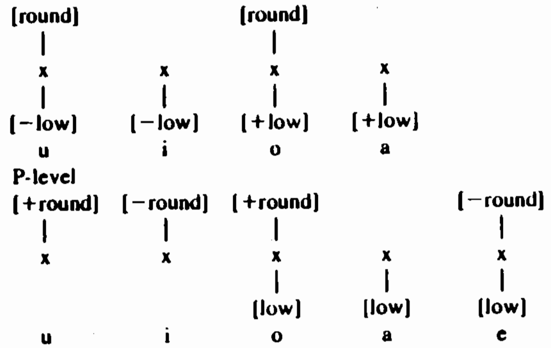
Syllable structure at the W-level in Yokuts is more varied, we observe, than at the P-level. While a long vowel is not permitted in a closed syllable at the P-level, it is permitted at the W-level. This is not an uncommon distribution; J. Leer has argued convincingly for this alignment of syllabification in Alutiiq (Leer 1988).

If the vowel feature specifications given in (22) are correct for M-level and W-level, they certainly cannot be correct for P-level, which has the canonical five-vowel system {i,u,e,o,a}. I have argued (Goldsmith 1985; 1988; 1990, 300) that the canonical five-vowel system is the phonological result of a two-feature system in which one feature (essentially height, or aperture) is privative and the other (essentially rounding or backing) is equipollent (i.e., bivalent), as in (29). The vowel system consists of vowels with no fewer than one and no more than two associations.

(29)


The shift from the four-vowel system of (22) to that of (29), as in (30), is thus part of the interface 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 can 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/morphology interface, or a D-level/S-level relationship. In the interface sketched in (30), there is an inversion in the character of both features: Low is equipollent on one level, privative on the other, and the same holds for Round.

(30) M-level, W-level



The feature interfaces utilized in (30) are both interesting and straightforward, and devolve from the shift in feature type noted just above.

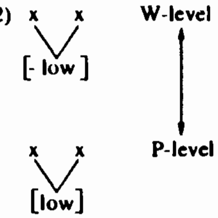
(i) Round is privative at M-,W-level, but equipollent at P-level; [+round] at P-level interfaces with [round] at W-level, while [−round] at P-level interfaces with Ø at W-level (as, of course, does Ø at P-level). These are the default correspondences that are imposed by the change in feature type.

(ii) Similarly, Low is equipollent at W-level and privative at P-level; again, [+low] at W-level interfaces with [low] at P-level, while [−low] at W-level interfaces with Ø at P-level. In general, featural shifts of this sort across levels follow the natural pattern given in (32). If M-,W-level vowels are subject to the level-particular restriction that the (equipollent) feature Low must be specified for each vowel, then no vowel will directly interface with the P-level, as desired. The lowering rule of (ii) 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 *i* or *u*; see (33).

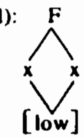
(31)

F: equipollent	F: privative
Level L_i	Level L_j
+F	F
−F	Ø

(32)



(33) Well-formedness Condition (P-level): where F is any feature.

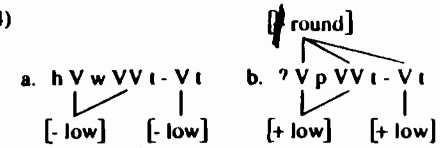


Rule:[13] Ø → [low]

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 thorough-going 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 three (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 pattern of stems as in (18c,d), with an "echo," or copied, vowel in the second syllable, is the result of autosegmental spreading. We may illustrate this as in (34), for the forms *hiweetit* and *ʔopootit*.

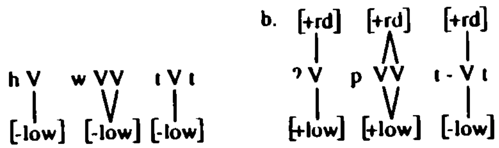
(34)



There is universal consensus among analyses of Yokuts 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: *hiwi:t, ṣudu:k, ʔopo:t, yawa:l,* etc.), and autosegmental theory requires (on one view, to which I subscribe; it strongly suggests, on another view) that the copying is done by autosegmental spreading illustrated in (34). However, the rule of Lowering only involves a single syllable in Yawelmani Yokuts (on other dialects, see Archangeli 1985); M-level *hiwi:tit* corresponds to P-level *hiwe:tit*, not, after all, *hewe:t-it* (or *hewe:tet*).

This suggests that the representation to which Lowering (24) applies is "syllabically local," as in (35), in a way in which the representations in (34) are not.

(35) a.



In the local representations of (35), each syllable has copies (so to speak) of the relevant autosegments (i.e., features).[14] While we have no specific additional grounds, we may speculate that this syntagmatic splitting of features is in some part due to the restructuring of the feature types (equipollent, privative) 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 intralevel 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 unsatisfactory remnant of the derivational conception has remained in our intralevel 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 in a new direction, and offer an explicit approach that addresses directly the aspect of the M/W/P harmonic model mentioned just above, the residue of the linear derivation within the intralevel 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 Chukchi mentioned above; cf. (4), (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 subparts 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 $n + 1$st rule will be possible without significant rule interaction arising. Composition of rules—guaranteed by our traditional notion of rule ordering[15]—is a simple way out of this technical morass of potential rule interaction and conflict.

In the previous section, I suggested that intralevel rules applied freely, with free reapplication. Such a suggestion naturally raises questions such as whether one could be certain that no instabilities would arise: whether, for example, a case might not arise in which a rule of epenthesis fed a rule of deletion, which in turn fed the epenthesis, which in turn fed the deletion, ad infinitum. One might equally attempt to demonstrate that free rule application of intralevel rules would not lead to unwanted and incorrect ambiguities of output, depending on which of two rules R_1 and R_2, 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 shrinking the conceptual 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 section, we take them to be models of the individual *levels* of the M/W/P model, along with the intralevel rules of the level in question. A single dynamic computational model, that is, represents simultaneously the representations, the tactics, and the intralevel 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.[18] A representation is viewed, then, as a state of the device, and its state—the representation—can be thought of as a point in a space of high dimensionality, a phase 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_n to representational state S_{n+1}. 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 system 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 activation from one unit to another, and to the specific arithmetic coefficients assigned to these 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 letting the stronger influence dominate. The final, or equilibrium, state of the device is the resultant 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 systems that are highly productive and regular in, no doubt, the vast majority of languages of the world, and second, these are systems in which rule interaction and long derivations have been the rule rather than the exception in discussions in the literature.

Our basic model is the same for the treatment of both accent and syllabification systems. We consider a linear sequence of units u_n, each of which has an activation value x_n which may be positive or negative. These units correspond roughly to the Row 0 units of the metrical grid[11] in the case of the accentual system, and to the skeletal tier in the case of the syllabification 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, to and sonority in the case of the syllabification operations on the skeletal tier) and *positional activation*—prominence assigned on the basis of position in the linear string.

Furthermore, each unit u_n is connected to its left- and right-hand neighbors in such a way that the more activated it is, the more it inhibits its neighbors, u_{n-1} and u_{n+1}. Conversely, a unit with negative activation excites its left- and right-hand neighbors; we refer to this relationship, therefore, as one of *polar inhibition*.

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

(36) $x_n^{t+1} = [\text{inherent activation of } u_n] + \alpha \cdot x_{n+1}^t + \beta \cdot x_{n-1}^t$

In general, the absolute value of the product of α and β must be less than about ~~2.0~~ for the system to be stable and achieve equilibrium; typical values, for example, are $\alpha = -0.5$ and $\beta = -0.1$, i.e., $\alpha < \beta$, though in the case of accentual systems this relation may be inverted. We refer to the final values x_n as the *derived activation* of the units u_n. 0.2

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 u_n, for whatever reason, has a positive inherent activation, then its left- and right-hand neighbors will have negative activation; their neighbors, in turn, will be positively activated; and the activation of unit u_n will spread out dynamically through the network. For example, if the fifth unit in a system with six units has an inherent activation of 1.0, with α and β as above, then the system will reach an equilibrium with the following values:

(37)

unit 1	unit 2	unit 3	unit 4	unit 5	unit 6
0.08	−0.16	0.31	−0.59	1.11	−0.11

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.[18] In addition, a threshold value (θ_s) 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 radiating outward 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 pre-antepenult, 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 competition 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.[19] Perfect Grid amounts, within this model, to any settings of α and β within (−1,0), and a low setting for θ_s, the threshold value. The resultant values can be calculated straightforwardly. If we assume that the values of (α,β) 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 syllable. In a system in which the local maxima are phonetically realized as stressed, the syllables that are marked with an acute accent will be stressed.

three-to six-

(38) (α,β) = (−0.5,−0.1)

3 syllables
o ó o
0.5 1.0 −0.1

4 syllables
ó o ó o
1.35 −0.7 1.13 −0.11

5 syllables
ó o o ó o
1.2 −0.39 0.54 1.0 −0.1

6 syllables
ó o ó o ó o
1.14 −0.27 0.32 −0.59 1.12 −0.12

While these are the values of α,β that appear to be correct in most of the systems we have looked at, the space of systems available is larger, and other equilibrium values are accessible with other settings of α,β (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 u_1 is a local maximum, and x_1 is greater than x_2; in region B, x_2 is greater than x_1, and u_2 dominates. What we see, then, is a system in which, by varying the explicit parameters, we can shift from one accent system to another, in the sense that within the distinct regions, distinct accent patterns are found for certain words or word types; but the general and crosslinguistic properties of accent systems, such as stress clash avoidance and alternating stress, are the result of the inherent architecture of the system.

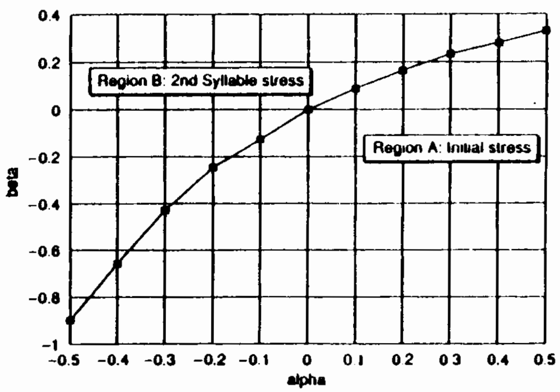


Figure 2.1

Any number of accentual conditions can be imposed simultaneously, and the output of the system is the quantitative resultant of these conditions. It is not hard to see that what we have achieved here, in effect, is something much like allowing rules to apply simultaneously and yet having a guarantee that there will be no unexpected and unresolvable problems of interaction due to composition; or rather, all problems of this sort

:table (and testable)

.

that even cyclic ac-
hion, by treating the
s simultaneous posi-
a language with, for
ı, the effect of cyclic
del by assigning ac-
syllable that is the
main is thus a matter
al stage in a deriva-
ı produce the correct

ificant results for the
: of those results, we
ıd indicate how this
d accentuation—not
nery for both, but in
stems are systems in
the system of accent

poken languages and
number of important
ıority have been pro-
nple, Steriade 1982,
Goldsmith and Lar-
tant is no doubt what
hich a segment s_i is a
ın that of its left- and
onority being greater
eing quantitatively a
ıe stressed elements
whose activation is a

.

ırtunities for a more
hat is incorrect about
ıe Sonority Principle,
ıterexamples where a
. . . , for example—is
llable nuclei and s_i is
:ns in languages with

what appear to be strong restrictions on the number of segments permitted in the onset and coda; such a case is discussed in detail in Dell and Elmedlaoui 1985, and reconsidered from the present point of view in Goldsmith and Larson 1990. In the latter paper, we suggest that the apparent violation of the Sonority Principle is due to the effects of the dynamic computational model, in the following sense.

The model consists of a sequence of units u_i as before, but with the difference now that each unit represents a segment rather than a syllable or mora; thus, the model as a whole represents the skeletal tier. The inherent activation of each unit is the sonority of the element (and we return below to how that quantitative measure is established, i.e., how it may be shown to be learnable). As with the accentual model described above, however, when we look for peaks, or local maxima, it is not peaks of *inherent* sonority that we seek, but rather of what we call *derived* sonority. Assuming values of α and β in the neighborhood of ~~0.5~~ and ~~0.1~~ as before, the high activation (or sonority) of the element s_{i+1} greatly decreases the activation of its left-hand neighbor, s_i, while s_i itself, greatly reduced in activation, affects the activation of its left-hand neighbor, s_{i-1}, very little; thus the derived activation of the sequence $s_{i-1}\ s_i\ s_{i+1}$ declines from s_{i-1} to s_i and then increases from s_i to s_{i+1}, and in general, the observed syllable nuclei may coincide with the derived sonority. -0.5 / -0.1

Larson has explored (in Larson 1990 and in prep.) the ways in which the salutary predictions of the Minimum Sonority Distance Principle (Steriade 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 sense, while additional results follow as well, as Larson (1990) discusses in connection with the syllabification pattern of Spanish. Without going into great detail here, some of the general characteristics 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,[20] syllables are stretches from one trough of derived sonority up to, but not including, the next trough of sonority.

Let us calculate when a sequence of segments $s_1\ s_2\ s_3$ may constitute an onset–nucleus sequence. That will be possible if the derived sonority (ds) of the segments satisfies the condition in (39).

(39) $ds(s_1) < ds(s_2) < ds(s_3)$

We would like to know what conditions this sets on the inherent sonority of the segments s_1, s_2, and s_3: which segments will be permitted to appear in such a sequence? Given any arbitrary values of α and β, the relation between inherent and derived sonority is a bit complex, but we can begin to get a feel for the effect if we consider as a first approximation the case where $\beta = 0$, and where we calculate only the first (but most significant) iteration in the step toward equilibrium. Under those conditions, the first inequality in (39) reduces to (40), where x_i is the inherent sonority of segment s_i.

(40) $(x_1 - \alpha x_2) < (x_2 - \alpha x_3)$, that is, $x_2 - x_1 > \alpha (x_3 - x_2)$

Thus, what this model suggests is not 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 describes 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 *...alba...* will be syllabified uncontroversially as *...al-ba...* in virtually all languages, the sequence *...abla...* is, as is well known, syllabified in some languages as *...ab-la...* and in others as *...a-bla....* The present model calculates the syllable division purely on the basis of the determination of the local minimum of derived sonority, and it is not difficult to observe that as we adjust the α parameter from 0 up toward 1.0, we observe the syllable division performed in these two different ways. With α at 0.0, *b* is a sonority minimum; as α increases, *l*'s derived sonority decreases, until it falls below *b*'s at a high enough value of α. Precisely where this point is depends on the assignment of inherent sonority values to the individual segments, a matter 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 contributions 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 segment—but how much? For every feature f_i there is a corresponding amount c_i, and the task of determining the sonority hierarchy for a given language is then equivalent to determining the values of c_i for all the features f_i. We may refer to the set of coefficients c_i 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 α and β (0.5, −0.1), and it begins (or rather, we begin) with purely random values of the set of coefficients c_i. It then calculates a sonority curve, with its maxima and minima, on the basis of these totally random sonority coefficients, and compares its guesses 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 its guess is correct, the coefficients are not changed. When the device guesses that a segment *s* is a peak but it is not, then it decreases by 10 percent the coefficients c_i that contributed to *s*'s sonority, and conversely, when it fails to identify a true sonority peak, it increases by 10 percent the coefficients c_i of the features that contribute to the sonority of that segment. In short, it acts like a perception.

For the first dozen words, the device does no better than close to chance 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 approximately 82 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 explicit 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 accentual system (we may say that an accentual unit a_i is associated with a skeletal unit s_j, or that they are associates, just in case s_j is a local maximum); (2) in what are called quantity-sensitive systems, the activation, or total amount of (derived) sonority, of a skeletal unit contributes to the activation of its associate on the accentual tier (Prince 1982).

In work in progress, Larson and I are exploring the nature of quantity-sensitive accentual systems (discussed in detail, for example, in Halle and Vergnaud 1987) by establishing local dynamic links between the local maxima on the skeletal tier and their associates on the accentual tier. In so doing, 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 (embedded in the model in the local connections of polar inhibition, as in (36)), and their propensity to assign greater accentual prominence to those syllables incorporating

.n general, then,
ɔe the activation
lar coefficient γ
ty-sensitive sys-
ɪit will be speci-
to one or more
ined by (36).

onsequences for
tionship between
onally been pos-
ɔroblem from the
ɔnological repre-
: application and
vel. In the latter
changes the con-
: conceptual dis-
pecific means by
dynamic fashion.
of recent work in
ɔ, and especially
icture of the dy-
.
ıre of phonology
structure, includ-
vill rearise as the
well-formedness
ɔring units. Such
nal well-formed-
els (and, presum-
ɪnd morphology),
ɜ levels.
this point, to be
y, however, is the
: notions of level
which were once
ackdrop of post-
ɪo more than one

Notes

1. This chapter has seen a number of versions circulated among colleagues over the past several years. It develops ideas that are discussed as well in Goldsmith (1989, 1990, in press a), and has been influenced by more linguists and nonlinguists than I could name. With regard to the notion of levels, I have been especially influenced by Charles Hockett and Noam Chomsky, as well as by J. R. Firth and Kenneth Pike; the reader will note the echoes of stratificational, prosodic, and tagmemic writing in some of the discussion herein. I have had a number of discussions with George Lakoff since 1987 on the notion of rule ordering and the challenges to intermediate representations. On the matter of harmonic rule application, the work by Singh and Sommerstein that I cite below deserves particular mention, as do in other ways lectures by John McCarthy at the 1987 Linguistic Institute at Stanford University. With regard to levels and harmonic application, various work on lexical phonology and phrase level phonology has been influential in that much 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 connectionism, as discussed, for example, in Rumelhart and McClelland et al. 1986, and I have found especially helpful the work of Smolensky (1986 and elsewhere), whose work on harmony theory was at the origin of my choice of name for the present work.

I have as well been influenced by the work of Sydney Lamb and of Charles Hockett, I should note. I am greatly indebted to my colleagues and students at the University of Chicago, especially to Jerry Sadock, whose work on autolexical theory stands behind the present discussion, and to discussions with the students in our phonology seminars, including Anna Bosch, Diane Brentari, Gary Larson, Karen Peterson, and Caroline Wiltshire. Geoffrey Huck, Younghee Na, and Jessie Pinkham have made very helpful comments on earlier drafts of this paper. This material was presented at the University of Toronto and the University of California at Berkeley, and I am grateful to the linguists there for helpful comments as well.

This material is based upon work supported by the National Science Foundation under Grant No. BNS 9009678.

2. Goldsmith 1990, chap. 5. On lexical phonology, see particularly Kiparsky 1982, 1985.

3. Perhaps the phrase is tendentious. There are no methodological principles given to us a priori that would allow us to know when a theory's constructs have strayed too far from what can be translated into a psychologically accurate model.

4. To my surprise (and dismay), I have found that these preliminary observations are occasionally greeted with a certain hostility among my otherwise eminently judicious and sensible colleagues. I wish to explicitly avoid giving the impression either that I am rejecting all of contemporary phonological thought (that is the farthest thing from my mind) or that I am suggesting that my proposal here is without precedent in the recent literature. On the contrary, it is rather obvi-

ous, as I have tried to suggest here and in Goldsmith (1990, chap. 6), 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 point of the present discussion is to encourage discussion, within the present context, of the relationship between derivations and levels.

A number of proposals in the current literature bear directly on the model discussed here. In certain respects, the "repair strategy" explorations of Singh (1987) and, following him, Paradis (1988), based explicitly on the proposals of Sommerstein (1974), are the closest, but in other respects their proposals are quite distinct (or perhaps "orthogonal" would be a better term) to the present discussion. Yip (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 noted. Some work of Alan Prince, of John McCarthy, and of Junko Itô is highly relevant, though the list of work that is highly relevant would be hard to complete once it was started.

5. 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 inclusively or exclusively as the reader wishes; my intention is that it be inclusive.

6. A recent and thoughtful discussion of these notions is found in Ladusaw 1988.

7. For purposes of concreteness, I take Hockett's position in Hockett 1961 to epitomize the theoretical position that I refer to here as American structuralism. Chomsky 1975 [1955] offers a fine explanation of the notion of level:

> The development of a system of phonemic representation does not conclude the process of linguistic analysis. We also want to discover the morphemes, words, and phrases of the language, and to determine principles of sentence construction 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 tokens, calling these systems "linguistic levels." A sentence token 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 token 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 1957:

> The central notion in linguistic theory is that of "linguistic level." A linguistic level, such as phonemics, 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 utterances. (P. 11)

8. 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 semanticists and the interpretive semanticists in the late 1960s derived from the use, by the generative semanticists, of the technical notion of derivation without a serious commitment to the view of syntactic analysis that lies behind its most natural use. Global rules (Lakoff 1969, 1970) seem like a formal abomination from the point of view of derivational syntactic analyses, though not from the point of view of a theory trying to link semantic and syntactic levels of analysis in as direct a fashion as possible—"direct" in the sense that an autonomous level of deep structure, distinct from semantic representation, was not included in the theory. This issue is discussed further in Goldsmith and Huck 1991.

9. A reviewer poses the question at his point as to whether we are to interpret the relation between levels as being achieved by rules acting simultaneously at each step, i.e., (MP,PM) and (PM,PT). The answer is yes, of course, though the structuralist would say that the word "simultaneously" 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+6 equals 8, and 8 equals 2 × 4, are we committed to saying that the two equalities are simultaneous when we write 2+6=8=2 × 4? In a word, no; to say they are true simultaneously would be misleading at best, a category mistake at worst.

10. András Kornai informs me that he has recently argued that such a position—developing a semantics, in effect, for phonology—has interesting consequences; I have not had the opportunity to study this work (Kornai 1991).

11. Chomsky 1975 [1955] provides a discussion of the introduction of representations that are motivated only by general simplicity of rule interactions (pp. 114–16 especially).

12. The present proposal leaves open a number of questions. If intralevel rules, for example, are unordered, is the claim not implicit that no cases will ever arise in which the order of application of two intralevel 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, adumbrated in the final section of this chapter. The essence of the approach is that what we take to be phonological rules can best be formalized as something approaching a field of force in a large-dimensional phase space; these fields can be arithmetically summed, and the sum is what determines the effect that the phonological ~~rule~~ undergoes.

13. This well-formedness condition must be interpreted as having the equivalent of a Structural Description and a Structural Change: the double association of a feature is the SD, and the [low] is what then must be there. More is to be said about this distinction.

14. Cf. Archangeli 1983 for a brief discussion and recognition of the problem that this material poses for a simple autosegmental account of Yawelmani Rounding Harmony and Lowering.

15. The guarantees that rule application will be kept well behaved from a formal point of view have been endangered by the increase found in the literature in appeals to "persistent" processes that are distinct from rules, persistent processes that regulate the representation by repairing structure or ensuring that it is well-formed (the first such example being the Well-formedness Condition of Goldsmith

1979 [1976]). The danger to be found in the increased use of such devices is this: Suppose we have an ordered set of rules, R_k, and a set of (more than one) persistent repair processes, P_j. After one of the rules (say, R_m) applies, all of the persistent repair processes ought to apply. But do they apply simultaneously 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 event, we will be left with two sets of ordered procedures, one set applying inside the derivation established by the other set. To my knowledge, none of these baroque consequences have been addressed in the literature, but the problems for current theory are very real. Itô (1986), in a major development in this line of research, is led to establish two persistent processes, Syllabification and Stray Erasure. A number of her examples, such as the discussion of Lardil, illustrate the requirement that after a (real) phonological rule has applied, Stray Erasure must not apply until *after* Syllabification has applied; 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 style of process interaction is impossible, nor even that it is reprehensible in any sort of way, but rather that familiar rule-oriented phonological theory has—on the basis of cogent and valid argumentation—backed itself into a corner where radical rethinking of rule organization is simply a necessity.

16. Or rather, many of these are phonological representations: not all states of the device need be defined as licit phonological representations.

17. Liberman and Prince 1977; Prince 1982; Halle and Vergnaud 1987.

18. This statement 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) neighbor. 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* (*extrapositionality*).

The statement in the text is also incomplete in that in our studies of stress systems, we have found it appropriate in a number of cases to take stress to be assigned to those 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 Scots Gaelic; see Bosch 1991; here syllables are defined as stretches from (but not including) one trough up to and including the next minimum.