Cognitive Phonology

George Lakoff
University of California at Berkeley

Presented at
The Berkeley Conference on Nondervational Phonology

May 26, 1989

Background

As a beginning student, I asked a question about generative phonology that students have asked many times:

Do people go through these step-by-step derivations in their heads every time they pronounce a sentence?

The answer was: No, of course not. So I asked the logical next question.

Then what makes derivations cognitively real if the intermediate stages never occur in the minds of speakers?

I got the performance-competence answer: The intermediate forms are real in terms of competence, but not performance.

Like other students of generative grammar, I soon learned to stop asking such questions, at least for a while. But with the advent of connectionist theories of mind, the questions have returned with a vengeance. Connectionist cognitive science suggests that the brain IS the mind and that anything that is cognitively real is represented in the brain. That includes all aspects of linguistic competence.

Neural processes occur in real time. Phonological derivations do not occur in real time, but in some abstract 'time' that cannot be put in correspondence with real time. Speakers process left-to-right, not top-to-bottom. There is no way that the intermediate stages of long phonological derivations of novel long sentences can be realized as such in a brain that functions in real time. All those intermediate stages of long derivations of sentences simply cannot be realized in the brain.

To me that suggests that there is something fundamentally wrong with the foundations of generative phonology, that all those rule orderings and cycles and principles are the product of a mistaken theory.

What I would like to do is suggest an alternative that will at once simplify phonology and also take a step toward bringing it into line with what the brain is capable of. One thing that connectionist models do naturally is characterize cross-dimensional correlations. Those of us working in cognitive
grammar have found that really complex syntax (of the 'non-core' variety, which is most of syntax) becomes tractable if it is formulated in terms of direct correlations (called 'grammatical constructions') between semantics/pragmatics and surface forms. I would like to suggest that phonology works in the same way cognitive grammar does, in terms of cross-dimensional correlations, where the relevant dimensions are levels of phonological representation.

Cognitive phonology is to be seen as an integral part of cognitive grammar. As such, it assumes that phonology, like the rest of language makes use of general cognitive mechanisms, such as cross-dimensional correlations.

**Cognitive Phonology**

Cognitive phonology characterizes correspondences between morphemes (as stored in the mind) and phonetic sequences. We will refer to these as the morphemic and phonetic 'levels'. In addition, we posit one intermediate level, the phonemic level, at which, among other things, constraints on word-level phonology are stated. I take these to be a minimal collection of necessary dimensions of phonological structure. I also think they are all that is necessary.

Generative rules are replaced by constructions, which state well-formedness constraints within levels and correlations across levels. In the default case, there is identity across levels. Cross-level constructions override such defaults. Other default cases, both language-specific and universal, are possible. Cross-level constructions are direction-neutral, and are intended to be used directly in either production or recognition. Constructions combine by superposition. That is, each construction imposes a set of constraints, and the constraints of the various constructions are simultaneously satisfied.

Cognitive phonology is set within a general autosegmental phonological framework. I assume that the ultimate formal framework in which this approach will be set will be a form of PDP connectionism, which is a simultaneous constraint satisfaction system.

My goal in this paper is modest: to illustrate how derivations, ordered rules, cycles, and application principles can be eliminated, at least for some common, but nontrivial textbook cases where it is claimed that they are necessary. I will also cover some classic cases where dialects are supposed to differ by rule ordering, as well as the case of Grimm's and Verner's Laws, which have often been seen as evidence for rule ordering.
Mohawk

Let us first consider an example from Mohawk that has been used to argue that phonological rules must be extrinsically ordered.

Mohawk has an epenthesis rule that, in generative phonology, would be stated as:

— Insert an /e/ between a consonant and a word-final glottal stop.

Epenthtetic /e/ interacts with two other rules. The first is stress assignment:

— Stress the penultimate vowel of the word.

In words whose final vowel is an epenthetic /e/, stress is antepenultimate. Hence it is claimed that:

Stress assignment precedes epenthesis.

The second rule that interacts with epenthesis is intervocalic voicing:

— Stop consonants are voiced between vowels.

This rule applies when the epenthetic /e/ is one of those vowels. Hence it is claimed that:

Epenthesis precedes intervocalic voicing.

Moreover, there is a vowel-deletion rule:

— A vowel before another vowel deletes.

When an underlying sequence ‘V C_0 V V C_0 #’ occurs, the second vowel deletes and the first vowel is stressed. This suggests the ordering:

Vowel deletion precedes stress assignment.

In addition, there is a vowel change rule:

— Stressed /ʌ/ becomes /ɔ/ before a nonanterior nasal consonant.

This implies:
Vowel change follows stress assignment.

Additionally, there is an ordering constraint that is probably an artifact of the analysis given. Underlying morphemes analyzed as beginning with /hr/ always simplify. The sequence /?+hr/ simplifies to /h/; elsewhere, /hr/ simplifies to /r/. This requires two ordered rules: cluster simplification followed by h-deletion.

A better analysis might be to analyze the supposed /hr/ sequences as underlying /r/ and have /?+r/ become /h/, which would eliminate one ordering constraint. However, for the sake of discussion, let us keep the traditional underlying /hr/ analysis.

Thus, the traditional generative account requires seven rules and six ordering constraints, with extrinsic ordering to a depth of four. Halle and Clements give a form which requires the application of six of these seven rules. The derivation has seven lines, with five intermediate forms.

\[
y e + \tilde{\Lambda} k + h r e k + ? #
\]

\[
y \tilde{\Lambda} k h r e k ? # \quad \text{(by vowel deletion)}
\]

\[
y \Lambda k h r e k ? # \quad \text{(by stress assignment)}
\]

\[
y \tilde{o} k h r e k ? # \quad \text{(by vowel change)}
\]

\[
y \tilde{o} k h r e k e ? # \quad \text{(by epenthesis)}
\]

\[
y \tilde{o} k h r e g e ? # \quad \text{(by voicing)}
\]

\[
y \tilde{o} k r e g e ? # \quad \text{(by h-deletion)}
\]

The cognitive phonology account is rather different. There are morphemic (M), phonemic (P), and phonetic (F) levels. (As a mnemonic, think of 'F' as standing for ‘fonetik.’) Instead of generative rules that that take input and give outputs, there are constructions, which are wellformedness conditions that hold within and across levels. Within-level constructions are preferred. Cross-level constructions state constraints that cannot be stated within a level. In a cross-level construction, the environment statement can occur at either level. The default situation is cross-level identity, and constructions override that default.

Each cross-level construction below has two levels of representations and a set of correspondence lines. A line with an 'X' through it indicates the absence of a corresponding element at a level. In a cross-level construction, there is an implicit implicational condition of the form ‘If ENV, then if X
then \( Y \), where 'ENV' is the environment of the rule, 'X' is the higher element in correspondence and 'Y' is the lower element.

Here are some sample Mohawk constructions.

**Vowel omission:** The first of two consecutive vowels at level M does not appear at level P.

\[
\text{M: } V + V \\
\times \\
\text{P: }$
\]

Stress: At P, a vowel that is penultimate is stressed.

\[
\text{P: } X_0 C_0 V_0 C_0 \#; \text{ If } X = V, \text{ then } X = [+\text{str}].
\]

We have stated stress placement as a one-level rule because it is possible to do so. Technically, it might have been stated as a cross-level rule, with the environment at M and the occurrence of [+str] at P, but the preference for minimizing cross-level constructions rules this out.

**Epenthesis:** When \( C \) precedes \( ?\# \) at level P, an /e/ absent at level P intervenes at level F.

\[
\text{P: } [] [ ] \\
\times \\
\text{F: } C \quad e \quad ? \quad #
\]

The P-level representation ' [ ] [ ]' simply indicates that the /C/ and /?/ are adjacent at P.

**Voicing:** At level F, intervocalic stops occur voiced.

\[
\text{F: } V \quad X \quad V; \text{ If } X = [+\text{obs}], \text{ then } X = [+\text{vd}].
\]
Vowel change: P-level \(\Lambda\) corresponds to F-level \(\mathcal{D}\).

\[
P: \Lambda \quad [-\text{syl}, \text{-ant, +nas}]
\]

\[
F: \mathcal{D}
\]

The two rules of h-deletion and cluster simplification necessitated by the analysis of underlying /hr/ are as follows:

H-deletion: An /h/ preceding an /r/ at P does not appear at F.

\[
P: h \quad r
\]

\[
F:
\]

Cluster simplification: A sequence /? h r/ at M appears as /h/ at P.

\[
M: ? \quad h \quad r
\]

\[
P:
\]

Here the ordering is essentially preserved, since cluster simplification is an M-P construction and h-deletion is an P-F construction. Another alternative would be to have both be M-P constructions and to let the connectionist equivalent of the elsewhere condition choose cluster simplification as a special case that take priority over the more general case of h-deletion. A third possibility is to substitute /r/ for /hr/ in the M-level representations and have a merging construction of the form:

Merging: M-level /? r/ corresponds to P-level /h/.

\[
M: ? \quad r
\]

\[
P: h
\]

For the sake of discussion, we will ignore the merging analysis.
The simultaneous satisfaction of the constraints stated by these constructions sanctions the following analysis. Instead of five intermediate levels, there is only one. The lines indicate sanctioned correspondences. The crossed-out lines indicate lack of an element at a level. And the rest is identical and is sanctioned by the default associations.

M: y e + Ê k + h r e k + ? #

P: y Ê k h r e k ? #

F: y Œ k r e g e ? #

To see just why the constructions given above are simultaneously satisfied by this analysis, let us unpack the analysis and check it out step by step.

—The sequence /e + Ê/ meets the M-condition of vowel deletion and the /e/ does not appear at P.

—The /Ê/ in P is penultimate, and requires stress as the stress construction states.

—The stressed /Ê/ in P corresponds to the /Œ/ in F, as sanctioned by vowel change.

—The /h/ in P precedes /r/ as indicated in /h/deletion, and so it does not appear in F.

—The /g/ in F is an obstruent between vowels, and since it is voiced it satisfies voicing.

—There is an epenthetic /e/ in F between /g/ and word-final /?/, which are adjacent in P. This satisfies epenthesis.

The moral: The Mohawk data do not in themselves show the need for ordered rules and derivations, since they can be accounted for with equal generality using cognitive phonology. What this shows is that ordered rules and derivations may well be an epiphenomenon of assuming the existence of generative rules that apply in order and manipulate symbols one step at a time.
Lardil

To illustrate a few of the things cognitive phonology can do, consider Hale's classic Lardil example, as discussed in the Kenstowicz-Kisseberth text.

An underlying form like /#tjumputjumpu#/ is pronounced as /#tjmp#utju#/ as a result of three rules, which in generative phonology would be stated as:

Apocope: V --> φ/ V C₁ V C₁ - #

Cluster simplification: C --> φ/ C - #

Nonapical deletion: [-syl,-apic] --> φ/ _ #

These rules produce the derivation:

#tjumputjumpu#

#tjumputjump# (by apocope)

#tjumputjum# (by cluster simplification)

#tjumputju# (by nonapical deletion)

Each rule applies in word-final position. In generative phonology they apply in sequence, producing new word-final positions as they apply, and hence create environments for subsequent rules. This is taken as evidence for rule ordering. Moreover, if the rules are stated as given and are not ordered, they will keep applying. For example, apocope could reapply to /#tjumputju#/ to yield /#tjumputj#/ which would be a wrong result. This can be avoided if the rules are ordered with apocope before cluster simplification before nonapical deletion.

Such cases are straightforward in cognitive phonology, and require only two levels. Let us assume they are the M- and P-levels. The rules are stated as follows:
Apocope:

M: \[ V \ C_1 \ V \ C_1 \ V \ # \]

P:

Cluster Simplification:

M: \[ C \ C \]

P: \[
\]

Nonapical Deletion:

M: \[-syl,-apic\]

P: \[
\]

Note that apocope is stated with the word boundary at level M, while the other two constructions have the word boundary at level P. In those constructions, the lines with X's through them terminate to the left of the word boundary at level P. What this means is that the M-level segment's corresponding (but unfilled) timing slot at the P-level falls within the word.

These constructions sanction the following pairing:

M: \# t j u m p u t j u m p u \# /

P: \# t j u m p u t j u \# /

In each case, the X'ed lines terminate to the left of the P-level word boundary, and there is nothing between that termination point and the word boundary. Thus, the three constructions can be satisfied simultaneously, with intermediate steps.

-The sequence /utjumpu#/ meets the M condition of apocope and the final /u/ in M has no corresponding element in P.
—The sequence /mp/ meets the M condition of cluster simplification, /p/ has no correlate in P, and its X’d line terminates to the left of # and nothing intervenes.

—The segment /m/ is [-syl,-apical], it has no correlate in P, and its X’d line terminates to the left of # and nothing intervenes.

Everything else is the same.

Note also that the following pairing is not sanctioned:

M: /# t j u m p u t j u m p u # /

P: /# t j u m p u t j # /

The second /u/ from the end in M is not immediately followed by a word boundary, and so the M-level condition on apocope is not met. Therefore, the absence of that /u/ at level P is not sanctioned. As in the case of generative phonology, each construction has a word-final environment, but the environments are not all on the same level.

Lardil has another rule that, in generative phonology, must be ordered with respect to nonapical deletion. It is a rule that lowers final vowels:

V --> [+low] #

The rule derives /kata#/ from /katu#/ and /pape#/ from /papi#/. It must be ordered before nonapical deletion. The reason is that underlying /naiuk/ undergoes nonapical deletion, resulting in /nalu/. If lowering applied after nonapical deletion, /nalu/ would incorrectly become /nala/.

Again, this is no problem in cognitive phonology. Lowering can be stated as:

M: V #

P: [+low]

This gives the correct result. The following analysis is sanctioned.
M: n a l u k #

P: n a l u #

The sequence / k #/ meets the M-condition for nonapical deletion and the /k/ is absent at P. But the /u/ in M does not immediately precede a word boundary and does not meet the M-condition for lowering to /a/.

Iteration

Rule iteration is a natural consequence of the theory of cognitive phonology. In generative phonology, certain rules must be specified as applying to their own outputs, either left-to-right or right-to-left. In cognitive phonology, iteration is an automatic consequence of the fact that a single construction may be simultaneously satisfied more than once in a word.

Let us compare the iterative shortening in Slovak and Gidabal. Kenstowicz and Kisseberth observe that both languages shorten a long vowel after a long vowel. Where they differ is in sequences of long vowels. Slovak shortens all but the first. Gidabal shortens every other one. Kenstowicz and Kisseberth suggest they have the same rule, namely,

\[ [+syl,+long] \rightarrow [-long] / [+syl,+long] \]

but they in different directions. Slovak, they suggest, applies the rule right-to-left to its own output, yielding derivations like:

V: C V: C V: C V:
V: C V: C V: C V
V: C V C V C V
V: C V C V C V

If the same rule is applied left-to-right in Gidabal, derivations like the following will occur, yielding alternate shortening.

V: C V: C V: C V:
V: C V C V: C V:
V: C V C V: C V
In cognitive phonology, this result can be achieved with no intermediate steps. The rules, however, will differ in the level at which their environment is stated. The Slovak rule will be stated as:

**Slovak Shortening**

M: [syl, +long] \( C_0 \) [syl, +long]

P: [-long]

This will sanction analyses like:

M: V: C V: C V: C V:

P: V: C V C V C V

The second, third, and fourth vowels in M will all meet the M-condition and their corresponding elements in P will all meet the P-condition.

**Gidabal Shortening**

M: [syl, +long]

P: [syl, +long] \( C_0 \) [-long]

This will sanction analyses like the following:

M: V: C V: C V: C V:

P: V: C V C V: C V

The second and fourth vowels in P will meet the P-condition of the construction, and their corresponding elements at the M-level will meet the M-level condition. As real time processing proceeds left-to-right (whether in production or perception), such pairings will be sanctioned.

The need, in generative phonology, for a right-to-left iteration in Slovak is particularly counterintuitive from the perspective of cognitive reality, since
real-time processing proceeds left-to-right and not the opposite. The cognitive phonology approach permits processing in real time left-to-right in both cases—without unnecessary intermediate stages.

Before proceeding to more complex cases, let us discuss two phenomena that are commonly given ad justification for rule ordering: dialect variation and historical change.

Dialect Variation

One of the main arguments for rule ordering has been the claim that dialects can differ minimally in their ordering of phonological rules. Bromberger and Halle (1989) cite Canadian dialects of English that have a rule that raises /ay/ to /Λy/ and /aw/ to /Λw/ before a voiceless consonant. They state the rule as:

Raising:

[-cons] → [-low] / [___, stressed] [-voiced]

This rule accounts for the pronunciation of “write” as [rΛyt], while “ride” is pronounced [rayd]. Similarly, “clout” is pronounced [kΛwɔt], while “cloud” is pronounced [klawd].

Dialects that have this rule may differ in whether the rule has an effect in environments where /t/ and /d/ merge to form a flap. The flap rule applies to merge /t/ and /d/ so that pairs like “plotting-plodding” and “wetting-wedding” are no longer distinguished by the voicing of their medial stop consonants. Bromberger and Halle state the flap rule as follows:

Flap:

[-cont, COR] → [+voi] / [-cons, stressed] ___ [-cons, unstressed]

They then observe that dialects containing both rules can differ in a way that can be observed by a difference in the ordering of these two rules. Dialect A (which they describe as having raising ordered before flap) has pairs like [rayDing] “riding” versus [rAyDing] “writing” and [kΛwD–Id] “clouded” versus [kΛwD–Id] “clouted.” In dialect B, which they describe as having raising ordered after flap, there is no raising in the voiceless cases; both “riding” and “writing” are pronounced [rayDing] and both “clouded” and “clouted” are pronounced [kΛwD–Id].

This dialect difference can, of course, be described readily with no rule ordering. Suppose we write S&H’s flap rule as:
Flap:

\[
P: \quad [-\text{cont}, \text{COR}]
\]

\[
F: [-\text{cons, stressed}] \quad [+\text{voi}] \quad [-\text{cons, unstressed}]
\]

The dialect difference can then be characterized by whether the environment of the raising rule is at the P-level or the F-level. In Dialect A, the environment is at the P-level:

Raising (dialect A):

\[
P: [-\text{cons, stressed}] \quad [-\text{voiced}]
\]

\[
F: \quad [-\text{low}]
\]

Since the \(d/-t/\) distinction is maintained at the P-level (in both dialects) this construction will yield the \(\text{klawD-D}-[\text{klAwD-Dd}]\) distinction of dialect A. In dialect B, on the other hand, the environment of raising is at the F-level.

Raising (dialect B):

\[
P: [-\text{cons, stressed}]
\]

\[
F: \quad [-\text{low}] \quad [-\text{voiced}]
\]

Thus, dialect B will distinguish “write” [\text{rAyt}] versus “ride” [\text{rayd}], but merge “writing” and “riding” to [\text{rayDing}]; the raising rule will not apply to “writing” because the voiceless stop is not present in that word at the F-level, which is where the environment for the raising rule is stated.

Basque Dialects

A similar solution will account for a minimal distinction in the Basque dialects of Baxtan and Markina, as described by Hualde (1988). Hualde gives the following forms as representative of the difference between the dialects. The \(/-a/\) suffix is the absolutive singular marker.
Baztan | Markina
---|---
/egun-a/ | [egune] | [egune]
/gison-a/ | [gisona] | [gisone]
/buru-a/ | [burue] | [burue]
/aşto-a/ | [aştua] | [aştue]
/mendi-a/ | [mendie] | [mendise]
/etşe-a/ | [etşia] | [etşie]

Hualde states two rules, **Mid Vowel Raising** and **Vowel Assimilation**.

**Mid Vowel Raising:**

[-low] $\rightarrow$ [+high] / ___ V

**Vowel Assimilation:**

a $\rightarrow$ e / [V, +high] C_0 ___ #

Hualde observes that the dialect difference can be accounted for by the following orderings of these rules:

**Baztan:** Vowel Assimilation before Mid Vowel Raising

**Markina:** Mid Vowel Raising before Vowel Assimilation

Thus, in Baztan, the final /a/ in /aşto-a/ does not assimilate since assimilation precedes the raising of /o/ to /u/; this ordering yields [aştua]. But in Markina, the /o/ raises to /u/ first, which triggers the change by assimilation of /a/ to /e/, yielding [aştue].

This dialect difference can be handled straightforwardly without rule ordering in cognitive phonology. Both dialects have the same Mid Vowel Raising construction.
Mid Vowel Raising:

P: [V, -low] V
   |
F: [+high]

The dialects differ in where the level at which the environment condition is stated. In Baztan, the environment is stated at the P-level, whereas in Markina, it is stated at the F-level.

Vowel Assimilation (Baztan):

P: [V, +high] C₀ a #
   /    
F: e

Vowel Assimilation (Markina):

P:
   a
F: [V, +high] C₀ e #

Thus, the occurrence of final /-a/ of /ašto-a/ as [e] in Markina but not in Baztan is explained by the presence of environment [+high] at the F-level in Markina but not in Baztan. Since the underlying /o/ of /ašto-a/ occurs as the high vowel [u] only at the F-level in both dialects, that [u] can trigger assimilation only where [+high] is in the environment of the assimilation construction at the F-level, that is, only in Markina.

Examples like these from Canadian English and Basque are also important from the perspective of historical change. These dialectal variants are most likely brought about by a process of change in which an environment at one level moves to another level. My guess is that, in the typical situation, a phonetic environment moves to the phonemic level.
Grimm's and Verner's Laws

A standard argument for rule ordering comes from Grimm's and Verner's laws. Historically, Grimm's law preceded Verner's law. Bromberger and Halle (1989) argue that those laws must be stated as ordered rules within the phonology of Germanic. They state the laws as follows:

Grimm's Law (Part I):

[-cont, -voiced] --> [+cont] / [-obstruent]

Verner's Law:


Thus, voiceless continuants produced by Grimm's Law I are voiced after stressless vowels by Verner's Law.

In cognitive phonology, there is no need for rule ordering to account for the data adequately and with no additional complication. As Deidre Wheeler has observed (p.c.), the two laws would be stated as follows:

Grimm's Law I:

P: [-obstruent] [-cont, -voiced]

F: [+cont]

Verner's Law:

F: If [V, -stress] X and X = [+cont], then X = [+voiced].

Thus, Grimm’s Law I is a P-F construction, while Verner’s Law is a well-formedness condition at the F level. The results are the same as with ordered generative rules.

Bromberger and Halle also claim that Part II of Grimm’s Law has to be ordered after Part I. They state it as:
Grimm’s Law II:

[-cont, -asp] --> [-voiced]

B&H argue that this must follow Grimm’s Law I in Germanic phonology in order to prevent the voiceless stops produced by this rule from becoming continuants. In cognitive phonology, Grimm’s Law II can, like Grimm’s Law I, be stated as a P-F construction, and so the two parts of Grimm’s Law will not interact.

Grimm’s Law II:

P:  [-cont, -asp]
    |
F:  [-voiced]

My guess is that this is not an isolated example. I surmise that other arguments for rule ordering based on diachronic evidence will collapse once they are looked at from the perspective of cognitive phonology. Rule addition arguments should come out looking pretty much like this case, and change-of-ordering arguments should come out looking like the dialect variation cases above, where there is not a reordering of rules but a shift in the level at which the environment of a construction is stated.

Icelandic

Let us now turn to some more complex cases. I will begin with modern Icelandic, which has been analyzed at length by Anderson (1974) and Kiparsky (ms), among others.

Anderson attempted to deal with the complexities of Modern Icelandic by changing the theory of rule ordering, to introduce the notion of local ordering. Kiparsky has proposed instead six strictly ordered rules that apply in two cycles (one lexical and one postlexical) plus the strict cyclicity condition, which blocks certain applications (controlling where certain rules turn on), plus the strong domain hypothesis, which allows one to indicate where certain rules turn off.

I will limit my discussion to the most celebrated case in the Anderson and Kiparsky studies—that of the interaction of u-umlaut, syncope, syllabification, and u-epenthesis. All of the various rule-ordering apparatus proposed by Anderson and Kiparsky is needed just to handle these cases. Thus, with relatively simple examples, the basic issues can stated very clearly.

I will argue that the correct interactions of those rules can be characterized with full generality with no rule ordering, no cycles, and no special
principles. We will require, as before, a morphemic (M) level, a phonemic (P) level, and a phonetic (F) level. Constructions linking the M-P levels correspond roughly to rules of lexical phonology, and constructions linking the P-F levels correspond roughly to postlexical phonology.

As before, a given construction may have its environment statement at any level. What is particularly interesting about Icelandic—and what has caused so much trouble for generative phonologists—is that Icelandic has a construction whose environment can be at either of two levels. Since such statements are impossible in generative phonology, recourse must be made to rule ordering, cyclicity, and special principles.

Let us begin with u-uumlaut and syncope. In generative phonology, these are stated as generative rules, and one can feed the other.

u-uumlaut: $a \rightarrow ö / C_0 u$

syncope: $[+\text{syl}, -\text{str}] \rightarrow φ / C \_ D V ; D = [-\text{syl}, +\text{cor}, +\text{lax}]$

Compare the minimally differing underlying forms: bagg+ul+i and bagg+il+u. Both rules apply to these forms, but to do so, they must apply in different orders:

- bagg+ul+i
  - bøgguli (by u-uumlaut)
  - bøggli (by syncope)

- bagg+il+u
  - bagglu (by syncope)
  - bøgglu (by u-uumlaut)

In the first case, u-uumlaut feeds syncope, while in the second, syncope feeds u-uumlaut.

Moreover, one cannot just say that u-uumlaut can apply anywhere, since Icelandic also has a u-epenthesis rule, which cannot feed u-uumlaut:

U-epenthesis: $φ \rightarrow u$ before an unsyllabified $r$.

Thus, underlying

$dag+r$

becomes

$dagur$

but cannot undergo-uumlaut to become
U-epenthesis must therefore be ordered strictly after u-umlaut.

Kiparsky deals with this by noting that /+u/ and /+i/ are inflectional suffixes, while /+ul/ and /+il/ are derivational suffixes. He therefore posits a cycle that applies first to the root plus derivational suffixes and later to the entire word including the inflectional suffixes. U-umlaut and syncope are specified as lexical rules and can apply on the first pass of the cycle, while u-epenthesis is specified as postlexical and cannot apply on that cycle. Thus, u-umlaut applies to böglu on the second cycle, whereas it applies to böggli on the first cycle.

In cognitive phonology, the situation is much simpler.

—Syncope is an M-P construction, with environment at M.

—U-umlaut is an M-P construction, with its environment condition holding at either M or P.

—Syllabification is a within-a-level construction. It is a wellformedness condition holding at the P and F levels.

—U-epenthesis is an P-F construction.

These constructions can be stated as follows.

Syllabification: A syllable consists of an onset cluster, a vowel, plus an optional onset cluster.

\[ P,F: \sigma[ C_1 V (C_1)] \quad \text{(syllabification)} \]

U-epenthesis: An epenthetic /u/ appears at F before an /r/ that is unsyllabified at P.

\[ P: \quad []-\sigma \]

\[ F: \quad u \quad r \]
Syncope: An M-level vowel which is followed by a lax coronal consonant and a suffixal vowel does not appear at the P-level.

M: V D + V  \hspace{1cm} (D = [-syl, +cor, +lax])

P:

U-umlaut: An M-level /a/ corresponds to a P-level ö if it precedes a /u/ at either level.

M: a \hspace{1cm} C_0 u

P: ö

Note that u-umlaut generalizes its environment over two levels; that is, the correspondence between /a/ and /ö/ will hold whether the environment 'C_0 u' is at the M-level or the P-level. /böggli/ is sanctioned because the environment is met at the M-level. /bögglu/ is sanctioned because the environment is met at the P-level.

Thus, we have the following level-representations for the cases discussed so far.

M: bagg+ul+i  \hspace{1cm} (syncope + u-umlaut with env at M)

P: böggli

F: böggli

M: bagg+il+u  \hspace{1cm} (syncope + u-umlaut with env at P)

P: bögglu

F: bögglu
M: dag+r

P: dagr  \hspace{10pt} \text{(u-epenthesis)}

F: dagur

Similarly, the M-level form /alim+um/ will correspond directly to the P-level /ólnum/, with /ọ/ sanctioned by the occurrence of a following /u/ at the P-level. The M-level /i/, due to syncope, does not appear at the P-level.

M: alim+um

P: ólnum

F: ólnum

The generalization of syllabification to both P and F, automatically produces what has been called 'resyllabification' as a result of F-level constraints. In addition, the principle governing vowel length is stated as a well-formedness condition at the F-level:

Lengthening: Vowels in open syllables are long.

F: If œ[C₁ V], then V = +long.

The simultaneous satisfaction of syllabification and lengthening at level F accounts for the long vowel in /li:furs/.

M: /li:fr+s/

P: /[li]:rs/  \hspace{10pt} ([li:] is a syllable)

F: /li:][furs]/

The /u/ in /li:furs/ is sanctioned by u-epenthesis, since /r/ is unsyllabified at P. Because the /u/ is present at F, the syllable [furs] is possible, leaving [li:] as an open syllable, which is long because it is open. The occurrence at F of the syllabification /li:][furs]/ rather than /li:furs/ should be an automatic consequence of an appropriate connectionist framework, where harmony is maximized (that is, energy is minimized). Complete syllabification produces greater harmony than incomplete syllabification. (See discussion of harmony below.)

There is no ordering, no cycles, no special principles. The M-P constructions work like lexical rules, the P-F constructions work like postlexical rules. The generalization of the environment for u-umlaut to both levels gets the
effect of opposite rule orderings. Simultaneous satisfaction of constraints at
the F-level produces the interactions of F-level syllabification and lengthening.

This is, of course, an oversimplification of Icelandic phonology. A fuller
account would require extensions and changes. Let me indicate how one such
change might go. Anderson points out that the statement of u-umlaut given
above is an oversimplification. There are really two parts of the process:
umlaut of /a/ to /ð/ before C₀ u, and vowel reduction of /ð/ to /u/ in
unstressed positions. The presence of both rules leads to rule ordering
difficulties in generative phonology, as Anderson discusses at length. Take the
underlying form

fátndað+um

The second /a/ precedes a consonant plus /u/, and therefore changes to by
umlaut to

fátndoð+um.

This contains an unstressed /ð/ which changes to /u/ by vowel reduction,
yielding

fátnduð+um.

This now contains another /a/ before a /u/. Umlaut applies again to produce

fátnduð+um.

This remains that way since the /ð/ is stressed. In a generative phonology
framework, umlaut must apply twice, once before vowel reduction and once
after vowel reduction. Therefore, umlaut cannot be ordered either before or
after vowel reduction. It was such paradoxes that led Anderson to the theory
of local ordering.

Such cases are straightforward in cognitive phonology. First, revise the
umlaut rule as follows:

Umlaut: M-level /a/ corresponds to a nonlow, nonback vowel at level P, when
followed by a /u/ at either level.

\[
\begin{align*}
M: & \{+\text{syl}, +\text{low}, +\text{back}\} \\
& \{C₀ u\} \\
P: & \{-\text{low}, -\text{back}\}
\end{align*}
\]

Then state vowel reduction as a wellformedness condition on level P.

Vowel reduction: At level P, unstressed vowels that are nonlow and nonback
are high.

P: If \{+syl, -str, -low, -back\}, then [+high].
In the case of an unstressed vowel, the simultaneous satisfaction of these two constructions would result in a /u/ at level P. In the case of a stressed vowel, vowel reduction would be inapplicable and the result would be an /ö/.

These two constructions would sanction the direct correspondence:

M: fátnað+um
P: fótnuð+um.

Here, there are two occurrences of umlaut and one of vowel reduction, all satisfied simultaneously.

—The correspondence between /a/ at M and /ö/ at P is sanctioned because /u/ follows /ö/ at level P.

—The correspondence between /a/ at M and /u/ at P is sanctioned because /u/ is [-low,-back] and /a/ precedes an occurrence of /u/ at M. This occurrence of [+high] /u/ rather than [-high] /ö/ at P is sanctioned because the /u/ is unstressed.

In summary, these cases from Modern Icelandic, which are usually taken as showing the need for complex rule ordering apparatus, including cycles and special conditions, actually show no such thing once cognitive phonology is brought into the picture.

Yawelmani

Yawelmani is well-known as a language that, from the point of view of generative phonology, requires considerable rule ordering. We will consider six rules of Yawelmani, all of which must enter into ordering relations, and then discuss how such phenomena would be treated in cognitive phonology.

Let us begin with four rules that reside at the center of the complexity of Yawelmani phonology—epenthesis, harmony, lowering, and shortening—which in generative phonology must apply in that order. The key rule among these is vowel harmony:

Vowel harmony: Vowels become round and back when following a round vowel of the same height.


Examples include the suffix /hin/, which appears as /hun/ when following a /u/, and /al/, which appears as /ol/ when following an /o/.
Vowel harmony must be ordered after epenthesis, which inserts an /i/ to break up consonant clusters.

Epenthesis: $\phi \rightarrow i / C _ - C \{C, #\}$

One reason is that epenthetic /i/ undergoes harmony. Thus, we have the derivation:

/?ugn+hin/  (underlying)

/?uginhin/  (epenthesis)

/?ugunhin/  (harmony on second vowel)

/?ugunhun/  (harmony on third vowel)

Another reason is that an epenthetic /i/ can block harmony. For example, underlying /logw + xa/ does not undergo harmony to /logw+xo/. Instead /logw + xa/ undergoes epenthesis to /logi+w+xa/, which introduces an /i/ that blocks harmony of the /a/.

Harmony must also be ordered before the rule of vowel lowering, which lowers long vowels /u:/ and /i:/ to /o:/ and /e:/.

Lowering: [+syl,+long] $\rightarrow$ [-high]

Those occurrences of /o:/ and /e:/ which derive from underlying /u:/ and /i:/ act like /u:/ and /i:/ with respect to harmony. For example, underlying /o:/ will trigger harmony, as when /do:x+al/ harmonizes to /do:xol/, while underlying /c'u:m+al/ does not undergo harmony, but then undergoes lowering of /u:/ to /o:/ to yield /c'o:mal/, which violates harmony at the phonetic level. Without a lowering rule ordered after harmony, we would expect the nonoccurring /c'omol/.

Then there is vowel shortening, which shortens vowels before two consonants.

Shortening: V $\rightarrow$ [-long] / _ C C

Thus, underlying /sa:p+al/ retains its long vowel, but /sa:p+hin/ shortens to /saphin/. Shortening must follow lowering, to account for those cases where long /u:/ first triggers harmony, then lowers (only long vowels lower), and then shortens. Here's an example:
s u d u: k + h i n (underlying)
s u d u: k + h u n (harmony)
s u d o: k + h u n (lowering)
s u d o k + h u n (shortening)

Shortening must also follow epenthesis, as we see from the contrast between /?amlal/ and /a:milhin/. Underlying /?a:ml+al/ has a long vowel /a:/ that occurs before two consonants /ml/. The /a:/ therefore shortens to /a/, yielding /?amlal/. But underlying /?a:ml+hin/ has three consecutive consonants /mlh/ which triggers epenthesis yielding /?a:milhin/. This breaks up the /ml/ cluster after the long vowel /a:/, which prevents shortening of /a:/ to /a:/—but only if shortening is ordered after epenthesis. The reverse order would predict the nonoccurring /?amilhin/. As a result, we have rule ordering to a depth of four:

epenthesis < harmony < lowering < shortening

Before introducing further complexities, let us how cognitive phonology can handle such a situation.

The most obvious part of the analysis is that epenthesis must be an M-P construction, while shortening in an P-F construction.

Epenthesis (first approximation):

M:  C  C  {C,#}
     / * /
     
P:  []  i  []

The square brackets at the F-level indicate the placement of /i/ at the F-level relative to the corresponding M-level consonants.

Shortening:

P:  [+syl,+long]  C  C
     
F:  [-long]
These two constructions would, obviously sanction the analyses:

M: ? a: m l + h i n
   (epenthesis)
P: ? a: m l + h i n
F: ? a: m l + h i n
M: ? a: m l + a l
P: ? a: m l + a l
   (shortening)
F: ? a m l + a l

Having a distinction between M-P and P-F constructions is equivalent, for such cases, to allowing one instance of rule ordering. Let us now turn to the interesting cases.

In order to correctly characterize the interaction between harmony and lowering, lowering must be a P-F construction and harmony must be a well-formedness condition at P.

Lowering: Long vowels at P are nonhigh at F.

P: [+syl,+long]
   |
F: [-high]

Harmony: At level P, every vowel that follows a rounded vowel with the same value of the feature [high] must be rounded and back.

P: If [+syl,+rd,a high] C X, then if X = [+syl,a high], then X = [+rd,+back].

Since harmony is a P-construction and lowering is a P-F construction, the M and P levels will be the same in cases like:

M: c' u: m + al
P: c' u: m + al
   (lowering)
F: c' o: m + al

- The harmony condition is irrelevant since /u:/ and /a/ have different values.
for the feature [high] at P. The lowering construction links /u:/ at P and /o:/ at F.

Harmony will however have an effect in the following case:

M: do: s + a l
P: do: s + o l
F: do: s + o l

—The harmony constraint is met at P.

—The link between /a/ at M and the corresponding /o/ at P is sanctioned for the following reason: /a/ and /o/ have the same features, except for those required by the harmony constraint. In particular, they have the same value of [high].

What is of interest here is that the difference between /a/ at M and /o/ and P is sanctioned not by a cross-level construction, but by the P-level harmony constraint plus the default condition of cross-level identity for everything not differentiated by the fitting of constructions.

The role of default identity is also crucial in the case of the interaction between harmony, lowering, and shortening.

M: s u d u: k + h i n

P: s u d u: k + h u n (harmony)

(lowering and shortening)

F: s u d o k + h u n

—The last /u/ at P is in harmony with with the previous /u/. The /i/ in M has the same features as the /u/ in P, except for those required by vowel harmony.

The lowering and shortening constructions are simultaneously satisfied by the P and F level representations. We can see this by superimposing the lowering and shortening constructions. Jointly they say that a long /u/ at P preceding two consonants corresponds to a short /o/ at F. This is just the relationship between the P and F levels, with everything else kept the same. Note that default identity does not apply construction-by-construction. Instead it takes into account the effect of all the constructions.

A technical aside: Within an appropriate connectionist framework, default identity would ideally be characterized in terms of Paul Smolensky’s
harmony theory (no connection to the term 'vowel harmony'). In connectionist theory, highest harmony corresponds to 'least energy.' I conceive of constructions as increasing harmony (in Smolensky's sense) both within and across dimensions. Default identity can then be stated as follows:

Identity across all levels is maximally harmonic except where harmony is increased by the fitting of constructions.

In a connectionist framework, cross-dimensional mappings always follow the path of greatest harmony. Default identity should be simply an automatic consequence of a connectionist framework in which constructions increase harmony.

Let us now return to our discussion of vowel harmony. The following example shows two interesting facets of cognitive phonology. First, vowel harmony, stated as a single-level construction, is automatically iterative. Second, it demonstrates something important about the role of markedness in cognitive phonology.

Let us return momentarily and reexamine epenthesis. The epenthesis construction stated above can be seen to be a conflation of two separate principles: one characterizes the occurrence of an epenthetic high vowel and the other specifies that the unmarked high vowel is /i/. Thus, we could have stated epenthesis more simply as:

Epenthesis: When a sequence /C C {C, #}/ occurs at M, a high vowel occurs at P between the corresponding first and second consonants.

\[
\begin{array}{c}
M: & C & \star & C & \{C, #\} \\
\end{array}
\]

\[
\begin{array}{c}
P: [ & [\text{+syl, +high} ] & [ ]
\end{array}
\]

High vowel markedness (DEFAULT):

If [\text{+syl, +high}], then [-back].

Since markedness principles are required independently, this analysis simplifies epenthesis.

In any theory, default principles are such that they can be overridden by nondefault principles. In cognitive phonology, that can be accomplished by specifying that

(1) Default principles increase harmony.

(2) Nondefault principles increase harmony more than default principles.

Since harmony is maximized in any cross-dimensional mapping, nondefault principles will override defaults.
Given such a simplification of epenthesis and given this account of what defaults are, we can understand why epenthesis and vowel harmony interact the way they do. Consider underlying /?ugn+hin/. The epenthesis and vowel harmony interact to yield the following analysis:

\[ \text{M: } ? \ u \ g \ u \ n + \ h \ i \ n \ \ \ \text{(epenthesis)} \]
\[ \text{P: } ? \ u \ g \ u \ n + \ h \ u \ n \ \ \ \text{(harmony of second and third vowels)} \]
\[ \text{F: } ? \ u \ g \ u \ n + \ h \ u \ n \]

Instead of epenthetic /i/, there is an epenthetic /u/ at level P. This /u/, being a high vowel, is sanctioned by our simplified epenthesis rule. This /u/ is also sanctioned by vowel harmony, since it has the same value for the feature [high] as the previous vowel (/u/) and is back and round. Since vowel harmony overrides the default markedness condition for high vowels, the epenthetic vowel, in this context, is /u/ not /i/.

Note that the vowel harmony construction fits twice here, once on the second vowel and once on the third vowel. The disparity in the third vowel between /i/ at M and /u/ at P is due to the joint effect of vowel harmony and default identity, which we discussed above.

Another technical note: The notion of default identity plus the distinction between default and non-default constructions is what enables this analysis to work. These are basic notions in cognitive phonology and they are natural notions for a connectionist framework. In such a framework, a construction is represented as a pattern of connection strengths. Since connection strengths have variable values, it is possible for some constructions to have higher overall connection strengths than others. This means that they will have higher harmony values. Thus, harmony value distinctions between constructions are natural ways of distinguishing default from nondefault constructions in theories with connectionist foundations.

The constructions discussed so far are the hard part of the characterization of Yawelmani phonology within cognitive phonology. The rest of what appear to be rule ordering phenomena can be handled in a straightforward way. To illustrate this, we will consider two more rules that interact with those considered above.

Vowel truncation, in generative phonology terms, deletes the first of two vowels:

\[ V \rightarrow \phi / V _{1} \]

It applies in derivations such as:
? i l i: + a l #
? i l i: + l # (truncation)
? i l e: + l # (lowering)
? i l e + l # (shortening)

Truncation must be ordered before shortening. If shortening were ordered first, its environment ( _ C #) would not be met. It is the prior application of truncation that makes shortening possible.

The cognitive phonology analysis is straightforward:

Truncation:

M: V V
   *

P:

M: ? i l i: + a l # (truncation)
P: ? i l i: + l # (lowering and shortening)
F: ? i l e + l #

Vowel drop deletes a final suffixal vowel:

V --> φ / V + C _ #

Vowel drop must be ordered before shortening, since it feeds it, as in the following example.

t a x a: + k’ a #
t a x a: + k’ # (vowel drop)
t a x a + k’ # (shortening)

In cognitive phonology, vowel drop is an M-P rule, which yields the same result.
Vowel drop:

M: \( V + C V \# \)

\[
\star
\]

P: 

M: \( t a x a: + k'a \# \)  (vowel drop)

P: \( t a x a: + k' \# \)  (shortening)

F: \( t a x a + k' \# \)

In summary, Yawelmani does not provide evidence for any levels beyond the morphemic, phonemic, and phonetic levels. Given M-P, P-F, and M-level constructions, together with a characterization of defaults, all the data that were previously thought to require rule ordering and lengthy derivations can be accounted for without them.

Summary

The need for derivations and rule ordering in phonology is an artifact of the classical theory of generative grammar, which assumed that grammars were symbol-manipulation devices. Such devices have to start somewhere and then proceed in a step-by-step fashion. They cannot "look ahead."

Cognitive grammars, by contrast, are not of this form. They contain constructions, not generative rules. Constructions state constraints on cross-dimensional correspondences. Such constraints are simultaneously satisfied in any sentence.

Cognitive phonology is a part of cognitive grammar. Phonological constructions state constraints across, as well as within, levels of the phonology. The levels are the minimum: morphemic, phonemic, and phonetic levels. Any adequate theory of phonology will need such levels. Because environments in constructions can be stated at any level, there is no need for rule ordering or cycles.

One of the attractions of cognitive phonology is that, as part of cognitive grammar, other aspects of the grammar are directly accessible. Since cognitive grammars characterize correlations across various dimensions of structure, correlations between the phonology and various aspects of syntax, semantics, and pragmatics are directly stateable.

Another major attraction is that the constructions that make up a grammar can be used directly in models of cognitive processing. Moreover, the same constructions can be used in generation and recognition, since there is
no directionality in their formulation. The rules of generative phonology cannot be used in models of cognitive processing at all—and they were not intended to be so used. And the intermediate stages of derivations in generative phonology could never occur as cognitively real representations in any processing model.

Cognitive phonology has other attractive features. First, connectionist foundations allow for a theory of defaults and default overrides in terms of levels of connection strengths. Another attraction is that the elsewhere principle is an automatic consequence of a connectionist framework. Within a connectionist framework, a combination of more specific and more general patterns can be learned only by having the more specific patterns have higher connection have higher connection strengths. Thus, the elsewhere condition is automatic, and amounts to a special case of defaults.

The characterization of vowel harmony as a well-formedness condition at a specific level is another attraction of cognitive phonology. In general, we would like all autosegmental spreading rules to be well-formedness conditions of this sort. Still another attraction is the elimination of left-to-right and right-to-left iteration.

I have not yet reevaluated the need for stress cycles, but from a preliminary look, it appears as though stress cycles can be replaced by constructions that mention syntactic structure.

One thing is for sure: Cognitive phonology is different. It drastically changes what can and cannot be done in phonology. Its possibilities have just begun to be investigated. But changes along these lines are necessary if we are to make phonology cognitively plausible.

My introductory linguistics teacher, Morris Halle, used to say that if you want to criticize a theory, you should come with a new one that does as well or better. With respect to the classic examples that have been used to motivate generative phonology—to motivate deduction-like derivations and rule ordering—we now have an alternative theory that can do better.
References


Acknowledgements

I would like to thank John Goldsmith and Larry Hyman for taking the time to discuss these issues with me at length and for giving be the benefit of their considerable knowledge. Without their support and enthusiasm, this project would not have gotten off the ground. My use of harmony theory derives from an idea of John Goldsmith’s.

I would also like to thank Paul Smolensky, David Touretzky, David Rumelhart, and the other faculty and and students of the 1988 Connectionist Summer School at Carnegie-Mellon University for augmenting my knowledge of connectionism and for serving as sounding boards for many of these ideas.

This research was supported in part by grants from the Sloan Foundation and the National Science Foundation to the University of California at Berkeley.

A Bibliographical Note

Discussions of cognitive grammar and the use of constructions in syntax can be found in the following works.


*Language* 64,3:501-538.
