Phonology, part 3: More on ordered rule systems

1 Bulgarian revisited

The rules for Bulgarian vowels that we arrived at in the course of our discussion in class were the following:

i. **Stress Assignment**

\[ V[\text{stress } - ] \rightarrow [\text{stress } +]/\#(C_m \begin{bmatrix} \text{syll} & + \\ \text{cons} & - \\ \text{stress} & - \end{bmatrix} C_k)_- \]

ii. **Vowel Raising**

\[ \begin{bmatrix} \text{stress } - \\ \text{low } + \end{bmatrix} \rightarrow [\text{low } -] \]

iii. **Consonant Devoicing**

\[ C[\text{voice } +] \rightarrow [\text{voice } -]/\neg[\text{voice } +] \text{ (applies anywhere it can)} \]

The first of these rules says the following, literally: add stress to a vowel which is the only vowel in the word (not taking the option indicated by the parentheses) or to the next vowel over to the right, if the first vowel is stressless. We can rephrase this in a way which shows it in a more intuitive, ‘natural’ light: in a word that does not already have stress in its phonemic form, stress the rightmost word. The second rule says, raise a stressless low vowel. Clearly, these rules have to be ordered, because you get different results depending on which of the rules you choose to apply to a given monosyllabic phonemic form. Consider, for example, the first first word, which takes the forms [glas], [gloš]. We know, because of the alternation in the vowel shape (a~o), that the underlying, ‘phonemic’ form of the vowel must be /a/, for the reason discussed in class, in connection with the phenomenon captured by the third rule, Consonant Devoicing: if in some cases we have a form x alternating with a form y in a particular phonological context, and in other cases we have x staying constant in exactly the same context, then it will be impossible to explain what’s going on if we assume that x is the phonemic form in both cases, because why would x do two quite different things under exactly the same conditions? The situation can be graphed this way:

<table>
<thead>
<tr>
<th>Environment 1</th>
<th>Environment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>x \rightarrow y</td>
<td>x \rightarrow x</td>
</tr>
</tbody>
</table>

Such a state of affairs would make no sense. If we were looking at three different environments, where x in one environment corresponded to x in a second environment but to y in a third, that of course would be
completely straightforward: it would simply mean that there was no rule applying to \( x \) which affected it in the second environment. But the situation depicted here is entirely different: \( x \) in one environment sometimes corresponds to \( y \) in one environment and at other times corresponds to itself, i.e., remains unchanged. The prospects for accounting for this sort of scenario are bleak, to say the least.

You might be thinking, at this point, that clearly such situations can’t really exist, and that we must be missing something, some factor which distinguishes between the two cases. And you’re right, but only if what have in mind doesn’t involve there being something different about the environments. Because, by assumption, the environments in this scenario are the same. And also by assumption, the \( x \) in Environment 2 really is the same in all measurable respects as the \( x \) in Environment 1. Those two assumptions do seem to fit Bulgarian final consonant voicing alternation quite well—remember that we looked around to see if there were other factors that might be involved, such as the nature of the stem vowel—but the answer seemed to be a clear negative: there’s nothing different in the environments in which the consonant corresponding to the final C of the stem is voiced, on the one hand, and voiceless on the other. Just consider the case of \([\text{lůk}]/[\text{lůk}]\), \([\text{můk}]/[\text{můg}]\), \([\text{líst}]/[\text{líst}]\): if it were the difference between a high front and a high back vowel that made the difference between the first (where is no final consonant voicing) and the second (where there is), then we shouldn’t find the third, since here we have a high front vowel but no voicing. Of course, one might say that the latter pair shows no stress shift, and that the voicing of the final C depends on there being a stress shift in addition, but then why do we find \([\text{lěf}]/[\text{lěva}]\), where we get voicing with no stress shift? Clearly, this line of attack on the problem isn’t going to get us very far.

So what could the difference be? And here, the suspicion that we’ve got the picture wrong in the above table can be cashed out, so to speak, by assuming that the difference doesn’t involve what’s going on in Environment 2, but rather what’s going on in Environment 1. Of course, if we look at those two environments, they might well look the same—but the crucial point is, the same environment might have different effects on different inputs to the rule. In other words, suppose that we don’t have the situation in the preceding table, but rather the situation in the following one:

<table>
<thead>
<tr>
<th>Environment 1</th>
<th>Environment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x ) ←− ( y )</td>
<td>( x ) ←− ( x )</td>
</tr>
</tbody>
</table>

Now things are much simpler: the reason for the alternation \( x/x \) on one case but \( x/y \) in the other is that in the latter case, it’s \( y \) that is the source of the \( x \), not the other way around. And all we have to do is assume that Environment 1 takes \( y \) to \( x \), but that in Environment 2, \( y \) is unaffected. On the other hand, Environment 1 has no effect on underlying \( x \), so it surfaces the same phonetically in both environments. Nothing could be simpler. In the case of final C (de)voicing, the point is that if we assume that the phoneme in the cases that change voicing really is voiced, but that it devoices in final position, whereas in the cases where the consonant is voiceless in both cases, nothing changes between the two environments so the final consonant is voiceless in both cases, then there is no longer a problem in accounting for the shift between voiced and voiceless versions of the final consonant.

(In fact, we can generalize the rule to the version given above, which says, devoice a consonant when it precedes any environment which is not specified as being voiced. This means that a voiced consonant preceding the silence at the end of a word will devoice, but it also means that a voiced consonant preceding a voiceless consonant will devoice. If we allow this rule to apply ‘at all times’—anytime such a situation arises—we’ll get the right answer, as I show below.)

And the same line of thinking makes it possible to get at the source of the Bulgarian stress shift. \( x \) and \( y \), after all, don’t have to refer to actual phonetic segments; they can refer to any kind of phonetic situation at all—the principle is the same. So we could frame the problem of the stress shift in the following terms:
Apply the rule of thumb we discovered in connection with final consonantal voicing—assume that when you have both \(x/y\) and \(x/x\) in Environment 1/Environment 2, where Environment 1 is the same in the two cases and Environment 2 is also the same in the two cases, then \(y\) is the underlying form, the mentally real one—and you’re immediately led to say that in the cases where there’s stress shift, it’s the \([\text{stress -}]\) version of the vowel which is the form of the phoneme. Now this is a little trickier than the consonant voicing case, because the environments aren’t quite the same: in the first case, the schwa bears stress, and in the second, it doesn’t. But if you ignore that for a second and just work with the environments as given in the preceding table, you’ll see that we’re led to say that the actual situation can be displayed as:

\[
\begin{array}{cccc}
\text{Env. 1: } C_m C_n & \text{Env. 2: } C_m C_n \bar{\alpha} \\
V & V \\
[\text{stress +}] & [\text{stress -}] \\
V & V \\
[\text{stress +}] & [\text{stress +}] \\
\end{array}
\]

In other words, an underlyingly unstressed vowel will get stress in the monosyllabic situation. But it won’t get stress in the bisyllabic situation, when the affix \(-\bar{\alpha}\) is attached. On the other hand, an underlying stressed vowel will have stress in both situations. So now we’ve figured out the hard part—what make stress shift in some cases but not in other? Answer: the stress rule will shift mentally unstressed vowels to stressed vowels in monosyllables. But that means that we have to state the stress rule so that it only does that in monosyllables. When a schwa is added, the stress rule will have to give us the stressed vowel that we see. And that means that we need to phrase the stress rule so that it has no effect on schwa when the root vowel is stressed. Those considerations lead directly to the stress rule as stated above: add stress to the leftmost vowel of an unstressed phonemic form of a word. If a mentally stressed version of the root exists, then the stress rule, given the way it’s stated, simply doesn’t apply, and stress surfaces where we see it, because it was part of the word—part of its representation in the lexicon—all along.

To get a feel for how these rules work in action, let’s look at a few derivations of phonetic forms from phonemic representations, corresponding to the mentally real picture of these words in the minds of native speakers of Bulgarian:
<table>
<thead>
<tr>
<th>Indefinite</th>
<th>Definite</th>
</tr>
</thead>
<tbody>
<tr>
<td>/glas/</td>
<td>/glas + a/</td>
</tr>
<tr>
<td>glás</td>
<td>glásá</td>
</tr>
<tr>
<td>glás</td>
<td>glásó</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>/med/</th>
<th>/med + a/</th>
</tr>
</thead>
<tbody>
<tr>
<td>méd</td>
<td>medá</td>
</tr>
<tr>
<td>méd</td>
<td>midá</td>
</tr>
<tr>
<td>mét</td>
<td>midá</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>/most/</th>
<th>/most + a/</th>
</tr>
</thead>
<tbody>
<tr>
<td>móst</td>
<td>mostá</td>
</tr>
<tr>
<td>móst</td>
<td>mustá</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>/znák/</th>
<th>/znák + a/</th>
</tr>
</thead>
<tbody>
<tr>
<td>znák</td>
<td>znákə</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>/lév/</th>
<th>/lév + a/</th>
</tr>
</thead>
<tbody>
<tr>
<td>léf</td>
<td>lévə</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>/grózd/</th>
<th>/grózd + a/</th>
</tr>
</thead>
<tbody>
<tr>
<td>grózt</td>
<td>grózdə</td>
</tr>
<tr>
<td>gróst</td>
<td>grózdə</td>
</tr>
</tbody>
</table>

To make sure you understand how the underlying forms interact with the rules, and why we get the pronunciations we do, it would be very useful for you to supply the rest of the derivations giving rise to the remaining data. Another useful move would be to try to work out a few derivations with the ordered rules reversed, and see what you come up with.

It should be noted that this is not the only way to implement the solution given. I’ve assumed that the suffix for definiteness is a. But, as was suggested in the discussion in class, suppose it’s actually á. We still need to have an account of when the stress stays on this vowel, as vs. when it appears on the main vowel, and that again will come down to a difference between inherent, ‘built-in’ stress on the one hand and lack of phonemic stress on the other. It looks at first as though this solution is no more costly than the one proposed above: if we have an underlyingly unstressed indefinite form X, the definite form will be X + á, and nothing further needs to be said, apparently. But appearances of deceptive here, for if we have a word of the form /CVC/, we will form a string /CVC+ á/, and then we will need a destressing rule, which applies to a syllable immediately following a stressed vowel. But that’s not enough—we’ll still need a rule which adds stress to an unstressed indefinite word prior to the operation of Vowel Raising, thus requiring two rules to do the job that the rule of Stress Assignment given can do with a single rule. General parsimony considerations thus suggest that the formulation given above is preferable. Again, a useful exercise would be the following: take the forms /lév/ and /med/ and form the indefinite and definite form of each of them. Doing so will give you a good sense of what’s at stake in this alternative version of the rule system for Bulgarian.
2 Klamath

To round out this discussion of rule systems, I want to show you how the use of ordered rules can covert cumbersome and arbitrary-looking descriptions which do not make use of ordering statements into models of descriptive economy. The relevant data are from Klamath, an extinct Penutian language of southern Oregon, in which we find the following data:

<table>
<thead>
<tr>
<th>Phonemic</th>
<th>Phonetic</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>honlina</td>
<td>holli:na</td>
<td>‘flies along the bank’</td>
</tr>
<tr>
<td>honl'ya</td>
<td>holhi</td>
<td>‘flies into’</td>
</tr>
<tr>
<td>honl'a:pa</td>
<td>l'a pa</td>
<td>‘flies into the fire’</td>
</tr>
<tr>
<td>pa:lha</td>
<td>pa:lha</td>
<td>‘dries on’</td>
</tr>
<tr>
<td>yalyalli</td>
<td>yalyal?i</td>
<td>‘clear’</td>
</tr>
</tbody>
</table>

Look carefully over this data, and try to make sense of what the large-scale patterns are, by looking at the phonemic representations and tracing just what changes are required to get you to the phonetic forms listed in the second column. The first example shows you that the sequence /nl/ is realized as a lengthened or ‘doubled’ l. The second indicates that the sequence nl˚ is realized as a sequence of l+h. The third represents a pattern whereby n followed by the laryngealized lateral sonorant l is pronounced as l+?. In the fourth, a sequence of voiced and voiceless lateral sonorant takes the phonetic form of a sequence l+h, and in the fifth, a sequence of l plus the laryngealized form of l is represented in pronunciation by l followed by glottal stop. These phonological changes can all be described in essentially just these terms, in a more economical notation, but with no difference in what’s being said: /nl/ → [ll], /nl˚ → [lh], nl˚ → [l?], /l{l} → [lh], /l{l} → [l?]. Just looking over these distributional statement, though, you should have the sense that we’re missing something. This is a clunky, stipulative setup no matter how you look at it.

The giveaway that all is not well is the fact that we clearly have a close association between nasals and laterals when these occur in sequence, with certain nasal+lateral clusters behaving the same as purely lateral clusters—but the statements just listed take no account of the fact that out of these five statement, there are two sets of two in each of which we have the same output. According to the five rules just listed, this is a coincidence. Yet given the facts that (i) nasal+lateral sequences of phonemes yield lateral+lateral phonetic forms and (ii) nasals plus particular laterals wind up yielding the same lateral + glottal segment sequences as combinations of lateral + those same respective laterals, can we not do better than simply listing the various phoneme sequences and their corresponding pronunciations, as though there were no systematic patterns?

At this point, my advice to you is to pick up a pad of paper and jot down an effort or two to come up with a better story, using the various tools that we’ve gradually developed in the course—particularly rule ordering. Take your time, and try to reduce the rule set to a smaller number, which accounts, in a principled way, for the patterns of parallel phonemic-to-phonetic replacements I’ve just mentioned. When you’ve reached the point where you either have got something you’re satisfied with, or are just stumped, read on.

Let’s start from the observation that /nl/ is pronounced [ll], as per the first descriptive rule listed above. Suppose we just say that nasals become laterals in the environment of following nasals. Without worrying too much about technical mechanics, we could interpret this statement to mean that n+l becomes l+l, n+l˚ becomes l+h, and n+l’ becomes l+l˚. Now, we don’t actually see the sequences l+l˚ or l+l’ anywhere in the phonetic output. But we know these sequences exist nonphonetically—look at the phonemic forms for ‘dries on’ and ‘clear’. If we take nasal+lateral sequences to these forms, their subsequent fate will be exactly the same as that of the lateral+lateral sequences in those two Klamath words respectively.
effect, if we turn two forms $X_1$ and $X_2$ into $Y_1$ and $Y_2$, which already exist independently, then whatever happens to the independent $Y$ forms will also be the eventual destiny of what started out as the $X$ forms.

Once you see that point—which is not actually a very deep or mysterious one, but which does require some kind of specific mechanism to implement it—then the rest of the solution is just a matter of execution. But the execution in this case depends on our ability to convert one kind of sequence of sound units to another kind which also has an independent origin, and then subject the latter to further operations which yield the same result regardless of where the input to these operations came from. We need some way to ‘feed’ the output of process to the operation of another process—and this is exactly what rule ordering allows us to do.

So let’s start by implementing the first part of this solution, whereby nasal+lateral sequences converge with lateral+lateral sequences and undergo all subsequent changes that apply to such lateral/lateral combinations. We do this in a straightforward way via the rule

**Lateral Assimilation:** $n \rightarrow l/$_[lateral +]$_$

I continue here to appeal to certain properties of sounds which hold across (sometimes quite large) phonetic groupings. The property [voice +], for example, covers quite a range of sounds in English, from vowels to voiced stops; the property [strident +], on the other hand, sets the relatively small group of fricatives [f,s,s] and their voiced counterpart (and matching affricates) off from all other sounds in English. I am going to assume that laterality corresponds to such a property. In English, only $[l]$ (and its nondistinct phonetically voiceless counterpart) are [lateral +], but in the languages of the Pacific Northwest, that’s anything but true; in most such languages we have at least $[l],[t],[ɬ],[ɮ]$ and $[ɬ']$ are parts of the phonetic inventory and likely to all represent different phonemes, and $[ɬ]$, the laryngealized labial sonorant, is quite common as well as a phonemically distinct sound category. So laterality plays a very big role in the structure of a least certain phonologies, and provides us with good justification for taking this property to be part of the universal set of phonological attributes which we can legitimately call on in describing properties of human languages.

But it is also clear that once I’ve grouped the distinctly ‘l-like’ sounds together as [lateral +], I also need to be able to split them more finely—to say what it is, for example, that differentiates the laryngealized lateral sonorant from, say, $[t]$, which we make take to be (as in Klamath) a voiceless version of $[l]$, at a more fundamental level. What distinguishes these two Klamath laterals from each other? The phonetic distinction of partially/gradually voiced sonorant ([l']) vs. voiceless fricative ([t]) doesn’t really work here, because there’s nothing about these descriptions which suggests that such sounds should be related to the particular sound sequences which they appear to ‘turn into’, i.e., $lP$ and $lh$. We need to ensure that sequences $ll'$ and $ll$ ($= l$) receive an analysis which makes sense of the fact the they become, respectively, $lP$ and $lh$. And the clue here is that both partial voicing and lack of voicing are connected to states of the same part of the vocal tract—the glottis—and that the segments $P$ and $h$ correspond to just those same states of the glottis: gradual/partial voicing, as in $l$, involves a partial obstruction to glottal opening, connected to the obstruction whose full form is the glottal stop, while lack of voicing is at the other end of the spectrum: the vocal cords so widely spread apart that they never get to vibrate. In a sense, what happens to $l$ when it follows another lateral is that it sheds its own laterality and becomes the segment corresponding to the same very closed state of the vocal cords that characterized its lateral ‘parent’. And an exactly parallel process occurs with the voiceless lateral, which becomes a glottal nonlateral with the same completely open glottis as its own precursor $[l]$—namely, the voiceless fricative $h$. In a sense, a single process is occurring here, which we can call Lateral Conversion: a lateral associated with a particular glottal secondary articulation is changed into the nonlateral which is maximally associated with that articulation. The exact way to do this involves technical issues that are beyond the scope of this course. But you can get an idea of the general approach if we start by defining a feature [constricted], indicating whether or not there is significant obstruction of
the glottis. We also invoke a second property, [spread], which is positive only for sounds associated with a complete separation of the vocal cords. Then the rules we need can be stated as follows:

\[
\begin{align*}
\text{C} & \begin{bmatrix}
\text{lateral} & + \\
\text{constricted} & + \\
\text{spread} & - \\
\end{bmatrix} \rightarrow \begin{bmatrix}
\text{lateral} & - \\
\text{constricted} & + \\
\text{spread} & - \\
\end{bmatrix} / [\text{lateral} +] \\
\text{C} & \begin{bmatrix}
\text{lateral} & + \\
\text{constricted} & - \\
\text{spread} & + \\
\end{bmatrix} \rightarrow \begin{bmatrix}
\text{lateral} & - \\
\text{constricted} & - \\
\text{spread} & + \\
\end{bmatrix} / [\text{lateral} +]
\end{align*}
\]

For purposes of this exercise, we can take the nonlateral, constricted and unspread consonant to be \( \text{?} \), and the nonlateral, unconstricted and spread consonant to be \( \text{h} \). And it may have occurred to you that these rules are similar enough that they ought to correspond to different ‘settings’ of a single rule. In fact, they do and such a rule can be stated, but again, the technology involved is a bit more than would be useful for us to pursue at this point. It’s enough to be aware that there is a single rule, statable using a somewhat jumped-up version of the notation we’ve been using, which we may call Lateral Conversion, whose effect is translate a laryngealized lateral into its nonlateral counterpart, a glottal stop, and a voiceless lateral into its nonlateral counterpart, the fricative \( \text{h} \), when these glottally implicated laterals occur after an ordinary lateral. Thus, the rule system of Klamath can now be stated as follows:

**Lateral Assimilation:** \( n \rightarrow l / [\text{lateral} +] > 
\]

\[
\begin{align*}
\text{C} & \begin{bmatrix}
\text{lateral} & + \\
\text{constricted} & + \\
\text{spread} & - \\
\end{bmatrix} \rightarrow \begin{bmatrix}
\text{lateral} & - \\
\text{constricted} & + \\
\text{spread} & - \\
\end{bmatrix} / [\text{lateral} +] \\
\text{C} & \begin{bmatrix}
\text{lateral} & + \\
\text{constricted} & - \\
\text{spread} & + \\
\end{bmatrix} \rightarrow \begin{bmatrix}
\text{lateral} & - \\
\text{constricted} & - \\
\text{spread} & + \\
\end{bmatrix} / [\text{lateral} +]
\end{align*}
\]

This system will yield derivations such as the following:

- **Lateral Assimilation:** \( /\text{honl'a}/ \rightarrow /\text{pa\l'a}/ \\
- **Lateral Conversion:** \( /\text{holl'a}/ \rightarrow /\text{pal'a}/ \\
- **Lateral Assimilation:** \( /\text{hol\l'a}/ \rightarrow /\text{pa\l'a}/ \\
- **Lateral Conversion:** \( /\text{hol\l'a}/ \rightarrow /\text{pa\l'a}/ \\

7
It is important to realize that Lateral Conversion, properly stated, will collapse these two clearly parallel rules into a single rule. Thus, we replace five separate statements with two ordered rules, and motivate simultaneously (i) the convergence of nominal+lateral with lateral+lateral combinations, on the one hand, and (ii) the particular nature of the ‘de-lateralization’ which befalls the second element in such combinations. It’s precisely because systems of ordered rules enable us to discover a principled, natural basis for much of what we observe in phonology that they constitute part of the core content of our picture of sound system and sound structure in human languages.