Linguistics 601 *Introduction to Linguistic Analysis*

Winter, 2010

Robert Levine

2 February

Phonology, part 1: Phonetic distribution, rules, and mental representation

1 What distribution tells us

Start with a very simple fact: English sonorants which immediately follow voiceless stops are themselves devoiced; everywhere else, they're voiced. So we find a distribution of sonorants which contains the following data:

greed	[gɹI^{y}d]	*creed*	[k̥ɹI^{y}d]
bray	[bɹɛ^{y}]	*pray*	[p̥ɹɛ^{y}]
dwindle	[dwIndl̩]	*twin*	[t̥wIn]

We see the same kind of pattern with fricatives preceding nasals: while the phonotactics of English preclude a [#$_{\sigma}$z+consonant] (i.e., [z] in initial position followed by any consonant at all) sequence, the permitted [s]+ nasal counterparts show this devoicing systematically: *small* [sm̥ɒl], *snake* [sn̥ɛ^{y}k]. (I should point out here that the vowel [ɒ], a very low back rounded vowel, isn't included in my chart in Phonetic Handout 3. I'm not sure how many people have it, and the unrounded vowel of *law* is confusing enough without adding another vowel to the system that I may be the only person in class who actually possesses.) The point is that this devoicing extends even to consonant which are, in fact, nonsyllabic vowels, and which we therefore would expect to be unfailingly voiced no matter what context they appeared in. Not only is that expectation flat-out wrong, but, even more remarkable, we don't even notice the fact. Not even one person out of a hundred would recognize immediately that the second sound in *dwindle* and the second sound in *twin* are crucially different from each other, in exactly the same way that [z] and [s] are. Why is this?

I've spoken in class of the contrary relationship between predictability and perceivability—basically, you don't notice what you can take for granted. This inverse relationship is slightly different from our relative lack of attention to, for example, people's actions when these actions are completely expected and normal for them; in the latter kind of situation, we notice the action and say, well, what else did you expect So-and-so to do? But in such cases, we do at least notice what So-and-so has done or said. In the case of the automatic devoicing of sonorants, however, we don't actually notice the devoicing; usually, it has to be pointed out to us, and even then, it's hard to actually hear the devoicing.

But there's another point that's of fundamental importance which emerges from what I've just said: notice how natural it sounds to you that I interpret the appearance of voiceless sonorants in the post-voiceless-stop environment as a case of *devoicing*. In contrast, imagine how strange you would have found it if I had suggested that the voiced liquids, nasals and semivowels that appear in all other environements was the result of the voicing of 'basically' unvoiced sounds. The facts are: in some places all of the sonorants are voiced, and in other places, they're voiceless. Those are the phonetic facts. But the consensus seems to be that, in effect, the voiceless examples of the sonorants are kind of beside the point—fundamentally, sonorants are voiced, through thick and thin; as MENTAL objects they're always voiced, regardless of whether the voicing is actually pronounced acoustically. This is very important; it illustrates in a particularly sharp way the difference between, on the one hand, the *phonetic* level of represenation, in which the physical facts—articulatory and acoustic—are all that count, and on the other hand, the *phonemic* level—the existence of

a psychologically separate domain in which speech sounds exist and combine with other such mentally real sounds, where measurable physical properties are in effect replaced by cognitive representations which are, in some sense, usually much simpler and more structured than the corresponding physical facts. But given the separation between these two domains, it's clear that there must be a certain 'translation' mechanism between the two: when we actually *say* something, we have to produce sounds, and when we hear what someone else has said, we have to map the physical string of sounds back into that mental representation. That's what the field of linguistics called phonology is all about: the relationship between the mental and physical aspects of speech sounds and the nature of the translation rules between the two.

The implications of these considerations for distributional patterning is clear: a rule determines how a single mentally real object is expressed, or manifested, or 'spelled out' in a different domain. In the case of the sonorants, for example, we have a kind of instruction set in our minds that dictates the following: Sonorant $\rightarrow$ voiceless/$\#_\sigma$ voiceless stop__. Again, we have to note the syllable boundary (the rule does not apply, for example, in a word like *shipwright*, where [p] precedes [ɹ], with the latter fully voiced, since [p] occurs at the end of the first syllable and [ɹ] begins the second, placing this sequence outside the scope of the rule). The fact that we need to note the syllable boundary is inherently interesting, because it gives us another fact about the mental side of speech sounds: syllable structure has to be part of the mental representation, or the rule mapping the psychologically real to the physically realized forms of speech sounds wouldn't be able to make reference to it.

What makes this picture slightly trickier than I've described it so far is the fact that physical reality is not entirely irrelevant to the mental sphere. Recall my example of [h] vs. [ŋ]: these do not occur in mutually exclusive environments, yet we have no sense at all that these are different manifestations of the same sound. If we look at the physical properties of these sounds, the discrepancies are clear: a voiced sonorant involving an oral closure and a uvular place of articulation, versus a voiceless continous sound, with no oral blockage. About the only thing they have in common is that both involve a continuous flow of sound, hardly the basis for a perceived identity. The fact is, phonology is a separate domain from phonetics, but there is a strong tendency for mental representations of speech sounds to reflect at least in part their physical properties. Nonetheless, this tendency is hardly universal, and we often see cases in which classes of sounds pattern together which seem to be physically quite different. An example is to be found in the following data:

Proto-Bantu Voiced Obstruants

βale	'two'	kiɣa	'eyebrow'
leme	'tongue'	ɣiɣɛ	'locust'
taβe	'twig'	kulu	'tortoise'
pala	'antelope'	oŋgo	'cooking pot'
kondɛ	'bean'	tɛndɛ	'palm tree'
zɔŋgɔ	'gall'	zala	'hunger'
βɛɣa	'monkey'	zɔɣu	'elephant'
βɛmbe	'pigeon'	βele	'body'
limo	'god, spirit'	lɛlu	'chin, beard'
kaŋga	'guinea fowl'	eɣi	'water'
ɣɔmbɛ	'cattle'	kiŋgɔ	'neck'
lelɔ	'fire'	nto	'person'

Stop reading for a moment and try to solve the problem of finding how the pattern of complementary distribution between [β,l,ɣ] and [b,d,g] respectively works (answer at the end of this handout, and you'll get a lot more out of this example if you work on it yourself first and try to come up with a neat solution, before going to the end to see what the story is). The upshot is that somewhere along the line, Proto-Bantu

'decided' that the fricative analogue of its voiced alveolar stop was to be a lateral, rather than a normal medial consonant, in contrast to how it treated its bilabial and velar obstruants. From the phonetic point of view, this representsa somewhat bizarre outcome; but while phonology may be constrained by phonetic fact, it rarely if error literally mirrors it. All this makes the actual 'doing' of phonology trickier, and therefore more interesting, than it would otherwise be.

A nice example of where physical facts intervene in the formation of strictly mentally represented rule systems involves the formation of English plurals (as well as possessives, third-person singular verb marking, and contracted forms of *is* and *has*). We have the following distributional facts:

singular	[dɔːg]	[k^{h}æt]	[wač]
plural	[dɔːgs]	[k^{h}æ ts]	[wačəz]

The facts are mostly simple: the plural marker is some kind of alveolar fricative, and takes on the voicing of the final sound in the singular form of the word to which it attaches. The one major exception to this generalization is the class of singular forms whose final segment itself involves an alveolar fricative kind of articulation; the pattern is exhibited not just with words ending in [č] as in the above example, but in *lease/leases, sneeze/sneezes* and *judge/judges*. And in particular, regardless of whether the words belonging to this class of examples end in a voiced or a voiceless sound, the material added has the same form: [əz]. If there *is* an actual break in the pattern, we can live with that, but we ought to try to see whether we can reinterpret the pattern reflected in the examples such as *watch/watched* above in a way that makes them consistent with the simpler pattern.

One sign that we'd be on the right track if we followed this strategy is that the form of the plural in these cases is actually consistent with the simpler pattern: [ə] is of course a voiced sound, so the appearance of [z] afterwards is consistent with the rule that a voiced form of the fricative appears after a voiced segment. The problem is that the [ə] isn't actually part of the singular form of the word: we have [wač] as the singular, not [wačə]. It's not like the case of *cheetah/cheetahs*, where the plural form of the word for the fastest member of the cat family consists of singular [čI^{y}ə] plus the plural marker, which we predict to be [z] for the same reason that we get *dogz*, *robes*, and *foes*. In the case of *cheetahs*, the singular form does legitimately possess a final vowel, voiced as usual, so that [z] is exactly what the simpler rule would lead us to predict. But if the [ə] in *watches* doesn come as part of the package with the singular, and if it represents an objectionable complication to regard it as part of the plural marking in *just* these cases, where does it come from?

Let's look for a moment at another set of facts: the form of words for certain kinds of measurement devices. We have

interfere	*interferometer*
spectr-	*spectrometer*
therm-	*thermometer*
gas	*gasometer*
ped-	*pedometer*

Such examples are easy to find, and what's interesting about them is that they are clearly made up of three separate parts: an English word or foreign stem indicating something about what it is that's being measured; the Greek stem *meter* (itself derived from Proto-Indo-European *me-), meaning 'measure', and an *o*, [ə] or [ʌ^{w}], depending on how the final form of the word is stressed. In none of the cases shown, nor in most other cases for that matter, is the -*o*- part of the stem to which *meter* is added, and it certainly isn't part of the Greek form from which we get what has become a kind of suffix in modern English. So what is the -*o*- doing there?

There are many such cases, many of them somewhat idiosyncratic. So consider the following pattern:

passage	*passenger*	'one who takes passage on some vehicle, as a traveller'
message	*messenger*	'one who carries a message to some recipient'
porridge	*porringer*	'a shallow bowl with a handle for eating porridge or soup'
harbor	*harbinger*	'one sent ahead to arrange harbor (in the sense of lodging or shelter) for someone in authority;

Looking at these pairs of related words, the first thing that should strike your mind is the question, where did that extra *n* come from?? This intrusive nasal began to appear in words of a certain general form with the addition of the *er* suffix meaning '-kind of being or thing' (as in: a messanger is someone who is the sort of entity associated with the bringing of a message). The *n* adds nothing to the combination noun+er; it just shows up, intrusively, between the meaningful parts, much as *-o-* does in the *thermometer* cases.

Suppose that the [ə] in the case of *watch/watches* and similar pairs shows up in the same way, but on a completely regular basis (note that the intrusion of *n* just illustrated is far from predictable; we have, e.g., *village* + *er* → *village*r, not **villanger*; the old alternate form for *suffragist* was *suffrager*, not **suffranger*, and so on). Furthermore, we can state exactly the conditions under which it regularly appears: when the plural marker is added to a singular form which ends in a sound which has the following characteristics—

- a highly compressed airflow producing a particularly turbulent, intensely harsh sound, and
- a point of articulation in which the front (tip or blade) of the tongue is raised above its 'neutral' position in the mouth—

—we find a [ə] between that sound and the following fricative marking the plural. These two characteristics are important enough that they have been paired with technical terms applying to all sounds that manifest them: the first is *stridency*, for obvious reasons, and the second is *coronality* (based on the contour of the roof of the mouth, similar to the rise of a crown in outline). One way of writing the combination of properties in question is therefore

$$\begin{bmatrix}\text{coronal} & + \\ \text{strident} & +\end{bmatrix}$$

This notation constitutes the description of a *class* of sounds, those which are positively specified for both properties. In English, the labiodental fricatives, the alveodental and alveopalatal fricatives, and the alveopalatal affricates are all [strident +], but, as you can prove to yourself fairly easily, articulating the labiodentals doesn't require you to raise the front part of the tongue in your mouth, which is what the [coronal +] part of the description requires. So what's left, when we look at *both* of these properties, is just the set of alveodental and alveopalatal fricatives, and alveopalatal affricates. And those are exactly the sounds which, when they appear as the final sound of the singular form, trigger the appearance of the 'intrusive' [ə] sound which separates the end of the singular form from the alveodental fricative marking the plural.

Armed with this new perspective on [ə] in the plural—that it literally comes out of nowhere to break up certain kinds of consonant sequences—we are in a position to simplify our account of the phonology of the plural fairly dramatically, by means of a rule of the form

$$\text{Schwa Insertion:}\quad \emptyset \rightarrow \text{ə} / \begin{bmatrix}\text{coronal} & + \\ \text{strident} & +\end{bmatrix} \_ \begin{bmatrix}\text{coronal} & + \\ \text{strident} & +\end{bmatrix}$$

The notation here is standard: $\emptyset$ denotes a string of length zero, so that $\emptyset \rightarrow \psi$ means, in effect, that whatever ψ represents is introduced, or inserted, by the rule in the context indicated. The underscore _ indicates the specific context in question.Thus, the rule should be read as follows:

> Schwa must appear between two adjacent phonetic segments which are both appropriately described as [coronal +, strident +]

An application of the rule will replace a string of two coronal, strident sounds with a sequence in which the original string is broken up by a schwa, which is exactly what we have in the case of the plural.

But a certain caution is necessary here: the precise formulation of the rest of the rule system is all-important. The whole point of Schwa Insertion is to create an environment in which the plural is expected to be [z], just as in the case of *dogs* and *cheetahs*. But this outcome implicitly depends on schwa already being in place when the 'decision' is made on whether the plural marker is to be voiced or voiceless. In other words, the insertion of schwa must *precede* the determination of voice for the fricative marking the plural, in order to have that determination yield the correct (voiced) result. We therefore need to order the rules involved, so that Schwa Insertion occurs first, and then the voicing rule.

The other question we need to decide is, just what *is* the rule determining voicing in the case of the plural (and the other morphological forms which obey the same rule)? We have three choices: we can assume that the phonemic form is /s/ and that voicing is added to this fricative to yield [z]; we can assume the opposite sequence (a mentally real, or 'underlying', /z/ is devoiced when following a voiceless final sound; or we can assume that the element which marks the plural is simply unspecified for voicing, and in chameleon fashion assumes the voicing 'coloration' of whatever precedes it. Any one of these three stories will work; most textbooks, so far as I can tell, assume that there's an underlying /z/ which is then devoiced in the appropriate environment. This solution actual allows a slightly simpler statement of the rule than the other two formulations, but whichever one we choose, it must—for the reason just noted—only occur *after* the insertion of ə. Thus, our final rule system has the following form:

$$\text{Schwa Insertion:}\quad \emptyset \rightarrow \text{ə} / \begin{bmatrix}\text{coronal } + \\ \text{strident } +\end{bmatrix} \_ \begin{bmatrix}\text{coronal } + \\ \text{strident } +\end{bmatrix} \; > \;$$

$$\text{Devoicing:}\quad \begin{bmatrix}\text{coronal } + \\ \text{strident } + \\ \text{voice } +\end{bmatrix} \rightarrow [\text{voice } -] / [\text{voice } -] \_$$

where > indicates an *ordering* relation. It is in fact characteristic of phonological rule systems that the rules belonging to these systems are at least partially ordered; we'll see many examples shortly. Note that the ordering here is nontrivial; if the reverse ordering applied, we would start with /watč+s/ and wind up with *[wačəs], since, on the scenario we're now entertaining, the devoicing would have applied *before* schwa was inserted.

The following derivations show how the rules presented will apply to yield the correct phonetic results on the basis of the phonemic input given between / / slashes:

	/kæt+z/	/dɔg+z/	/wač+z/
Schwa Insertion	kætz	dɔgz	wačəz
Devoicing	kæts	dɔgz	wačəz

These are exactly the forms that we observe, of course, and we have obtained them in a way which makes the relevant voicing rule completely general. The cost of this generality is the introduction of the schwa insertion rule, but in a sense, there is no real cost here, because positing such a rule ultimately allows us to establish

a still more general rule uniting the phonetic form of plurals and other morphological combinations with the phonetic form of the past tense (and other morphological combinations). Meanwhile, it's worth noting that in terms of communicative function, the schwa insertion rule dramatically increases the signal to noise ratio over what we would have to confront in the absence of such a rule. Strident sounds are, first and foremost, *noisy*, and this noice tends to obscure any immediately following part of the speech signal—particular when that phonetic followup is noisy itself! Any sequences such as [šs] or [ǰz] would, without question, seriously compromise the hearer's ability to discern the fricative marking the plural, and thus reduce the effectiveness of verbal communication in general. The intervention of [ə] in these contexts allows the physical signal carrying the information about plurality to have maximum audibility, and hence represents a rule with an obviously 'adaptive' value, especially from the hearer's point of view.

But it is important to understand that the ordering of phonological rules, as outlined above, does not reflect a sequence of events in real time. Rule ordering is a formal notion, not corresponding to an actual mentally realized procedure, for if it were, we would have the right to expect good evidence that it takes longer to produce plural forms when the plural suffix is added to roots ending in strident coronal sounds. There is in fact not the slightest evidence that such real-time delays occur. Rule ordering, rather, has to be understood as taking place in a kind of abstract, timeless realm of formal relationships amongst symbols, in much the same way that mathematical proofs correspond not to actual events in normal time, but to inherent relationships of priority which have a basis in logical inference that is quite independent of the forward flow of temporal order. We have to *check* the outcome of such derivations in real time, of course, but that's something completely different from the notion that those derivations themselves have an existence in real time.

Answer to the Proto-Bantu problem: The stop forms occur only following a nasal; the fricative forms and [l] occur in all other environments. This means that we have, in effect, three rules:

$$\left\{ \begin{array}{c} \text{b} \\ \text{d} \\ \text{g} \end{array} \right\} \longrightarrow \left\{ \begin{array}{c} \beta \\ \text{l} \\ \gamma \end{array} \right\} / \underline{\quad}$$

These rules could further be collapsed into a single rule—but *only* if we assume that the lateral resonant /l/ is actually the 'voiced fricative' version of the voiced stop /d/. This is a very strange treatment of a lateral liquid, however. It's possible that what we think of the voiced fricative analogue of [d], i.e., [z], was not present in Proto-Bantu—but we know that [s] was...