A Restricted Model of UR Discovery: Evidence from Lakhota

A fundamental tenet of generative phonology is that learners compare surface variants of morphemes, combining unpredictable information to establish URs that distinguish all surface forms. In this talk, I pursue the more restrictive hypothesis that learners base URs on a single surface alternant, even if this means that some unpredictable properties must be handled by mechanisms other than the UR (Hayes 1999; Albright et al. 2001). I present evidence for this approach from Lakhota. I also provide a computationally implemented algorithm that selects the optimal UR under this restriction, learns rules to derive the rest of the paradigm, and identifies exceptions, which must be lexically listed.

In Lakhota, some verbs display an invariant final vowel and final reduplication (1a), while others exhibit an [e]~[a] alternation and penultimate reduplication (1b):

(1) a. ‘good’ ‘spotted’ b. ‘fat’
   3sg. washte gleshka chepe
   3pl. washte-pi gleshka-pi chepe-pi
   REDUP. washte-shite gleshka-shka chep-chepe

In a traditional approach, learners could compare these forms and posit URs that distinguish them; following Boas & Deloria (1941) and Shaw (1980), we might say that ‘good’ ends in /e/ and ‘spotted’ in /a/, while in ‘fat’ the final [e]~[a] is absent underlyingly (/chep/), and is inserted by a later epenthesis rule. If learners are uncertain about the UR and must guess, there are many possible errors: they may assume that the final [e] of chepe is underlying and produce 3pl. *chepe-pi, that the final [e] of washte is epenthetic and produce 3pl. *washta-pi, and so on. Erroneous assumptions about the UR should always lead the learner to produce a “valid” paradigm, however; if the final vowel is underlying, the verb should behave as in (1a), and if not, as in (1b).

What we find in fact is that two additional, innovative paradigms have arisen: the first has [e]~[a] alternations but final reduplication: hansk/hanska-pil/hanska-shka ‘tall’. The second has invariant final [a] but penultimate reduplication: thokc/thokc-pil/thokc-thokc ‘different’. Neither of the new paradigms is compatible with any UR in the old system. I argue that these innovations make sense, however, if learners must choose a particular surface form (chepe or chepe) as the UR. In this circumstance, the best the learner can do is to choose the form that distinguishes the greatest number of words, which turns out to be the form found in the 3pl: /washte/, /gleshka/, and /chepep/. Under this analysis, the difference between gleshka and chepe is not reflected in their URs, but must be handled by an [a]→[e] raising rule for the singular. Thus, the [a] in gleshka must be learned as an exception, and the only error we expect is incorrect /al→[e] raising in gleshka-type words. Likewise, the difference between final and penultimate reduplication of /al-final roots (gleshka-shka vs. chep-chepe) cannot be because of an underlying difference in their final vowels, but must rather be the result of competing reduplication rules. The dominant pattern is penultimate reduplication; thus, we predict that learners may fail to learn that /gleshka/-type words have exceptional final reduplication, and extend penultimate reduplication to them. The restricted UR approach correctly predicts the two attested errors, and no others. Furthermore, because [e]~[a] alternations and reduplication patterns are learned as separate rules, with their own lists of exceptions, we can interpret the decoupling of these processes, which were once predictably linked.