The analysis of gradience in phonology: what are the right tools?

Bruce Hayes
UCLA

Gradient phenomena in phonology come in various kinds, of which two are discussed here:

- Gradient well-formed intuitions are found in **phonotactics**. Intermediate judgments usually arise for segment sequences that are rare in the language, as with English syllable onsets like [dw] or [bw].

- Gradience also arises when speakers must choose between **conflicting patterns of alternation**, in segmental phonology (Zuraw 2000) or phonologically conditioned affix choice (Albright 2002). We find a quantitative match between the statistics of the lexicon and speaker behavior, either in creating new forms or in rating forms proposed to them (Albright and Hayes 2003, Ernestus and Baayen 2003, Hayes and Londe 2006).

For the analyses of these phenomena, there is reason to go beyond traditional handcrafted grammars and instead seek grammars learned by machine-implemented algorithms. Since gradient intuitions are sensitive to the statistics of the lexicon, the best grammars are likely to be obtained by algorithms that can comb through the learning data, fine-tuning the grammar with greater care than humans can. Moreover, grammars learned by algorithm directly address a long-standing goal of generative theorizing (Chomsky and Halle 1965) namely to explain how acquisition is possible.

Thus, we need a formal framework for gradient grammar, coupled with a learning algorithm suited to learning the class of possible grammars under this framework. There are least two major current contenders for such a model.

- **Stochastic Optimality Theory** (e.g., Boersma and Hayes 2001) is based fundamentally on the same apparatus as standard OT. It locates constraints on a numerical scale that determines a probability distribution for how the constraints will be ranked whenever the grammar is invoked. From this it is possible to compute the probabilities of outputs, which can then be assessed against experimental or corpus data. One learning algorithm for Stochastic OT is the Gradual Learning Algorithm (Boersma 1997); others have since been proposed (Maslova in press, Lin, ms., Wilson, ms.).

- **Maximum entropy models** have a long and distinguished history in other sciences, but were first employed for phonology (as far as I know) only in 2003 by Goldwater and Johnson. In a maximum entropy grammar, the constraints are weighted, not ranked. Probability is assigned to candidates by a formula based on a weighted sum of their violations (for tutorial, see Hayes and Wilson, forthcoming, §3.2). The algorithms for learning constraint weights have been subjected to rigorous development (e.g. by Della Pietra et al. 1997); and they provably converge to the optimum set of weights, specifically, those which maximize the probability of the learning data under the constraint set.

In my talk I will try to make the following points:

- The well-worked out mathematics of maximum entropy, not surprisingly, yields a very high degree of precision when applied to data. This has important practical consequences for analysts. Another important advantage is that maximum entropy learning copes (in a rather surprising way) with the GLA non-convergence case pointed out by Pater (in press).

- Maximum entropy also offers a novel, and arguably more effective, approach to phonotactics. Unlike Stochastic OT, it can be used to assign a probability to every possible surface form, without regard to underlying forms. This circumvents a problem faced by previous approaches, namely the unconstrained search space that we encounter in trying to do phonotactics by finding all the forms that can derived from a rich base (Prince and Smolensky 1993, §9.1). Moreover, a maximum entropy grammar can match patterns in the lexicon very closely, and can project beyond the lexical data in modeling native speaker intuitions. I discuss an example from the current project of Hayes and Wilson (forthcoming) in modeling phonotactics with maximum entropy grammars.

- However, abandoning the rich base approach to phonotactics brings back an old challenge (Kisseberth 1970), that of unifying phonotactic analysis with the analysis of alternation—a problem often judged to have been solved by Optimality Theory. I will suggest that the OT account is not the only possible solution, and give an outline sketch of a “learning theoretic phonology” in which the connection between phonotactics and alternation is based on the course of learning, rather than the theory-internal connection posited by OT.