

*Mini-course at Brown University:*

## Embedding Grammar in a Quantitative Framework: Case Studies from Phonology and Metrics

Bruce Hayes

University of California, Los Angeles

May 18-22, 2009

An important research result of recent years is the discovery that speakers know much more about their language than classical generative grammars give them credit for. In particular, they possess detailed quantitative knowledge of the patterning of their language. This finding emerged as linguists tried shifting their research technique, from hand examination of isolated forms thought representative, to machine-aided searching of corpora.

Zuraw (2000), inspecting a Tagalog corpus for the well-known phonological rule of Nasal Substitution (/ŋp ŋb ŋt ŋd ŋk ŋg/ → [m m n n ŋ ŋ]) found that the rule is riddled with “exceptions”; i.e. words to which it does not apply. Yet the exceptions are statistically structured: in the aggregate, the rule applies more often when C2 is voiceless (/ptk/ > /bdg/), and also when it has a frontier place of articulation (/pb/ > /td/ > /kg/). These patterns are shown to be part of native speaker’s knowledge by experiments: in a “wug” test with novel stems, speakers behave probabilistically in a way that matches the lexical quantitative pattern. Classical generativist accounts, with their strict division into rules and exceptions, have no purchase on such patterns. Results similar to Zuraw’s have been obtained by several research groups, working in morphology and syntax as well as in phonology.

Such results lead some to suppose that linguistic theory has to be completely reinvented—replaced perhaps, by some kind of analogical system. To me, a more sensible strategy is to retain generative linguistics but make it more powerful by embedding the crucial elements of grammars (rules and constraints) into an appropriate quantitative framework. There is a good clue that this approach is the right one: constraints that influence quantitative patterning in one language often turn out to be exceptionless in another. For instance, the ban on voiceless consonants after nasals, which plays a quantitative role in Tagalog, is absolutely respected in the Yamato vocabulary stratum of Japanese. In sum: variation in language is patterned, and existing concepts of linguistic theory can be adapted to the characterization of such patterns.

The content of this mini-course will be a survey of how we might go about doing this. Possible quantitative frameworks to be examined will include stochastic optimality theory, maximum entropy, noisy harmonic grammar, and others. The empirical examples will largely be drawn from areas where I’ve done some work: English irregular past tenses, Hungarian vowel harmony, phonotactics, and the metrics of sung and written poetry. I will take a “consumer’s view” of the quantitative models, showing what they can and cannot do in service to the empirical research program. Throughout, I will keep an eye on the question of whether constraints have a basis in Universal Grammar, or whether they are discovered inductively.