

Learning Unbounded Stress Systems via Local Inference
Jeffrey Heinz
University of California, Los Angeles

This paper demonstrates how locality in phonology can be formalized to benefit learning. It is shown how an unsupervised batch learning algorithm learns the class of unbounded stress systems (Hayes 1981, 1995, Prince 1983, 1985, Idsardi 1992, Walker 2000, Bakovic 2004) using only a notion of locality—and no *a priori* Optimality-theoretic (OT) constraints (Prince and Smolensky 2004).

Hayes (2004) motivates a pure phonotactic learner by hypothesizing that a complete phonological grammar is more easily learned in stages: first the phonotactics of the target language is learned prior to the alternations. This strategy follows from the observation that children appear to acquire alternations later than the phonotactics, much of which appears to be in place by as early as one year. The learning mechanism described in Albright and Hayes (2002, 2003), for example, succeeds in part because it has been given phonotactic constraints that cue rule discovery.

Riggle (2004) has shown that if OT constraints can be written as finite state machines, then for most rankings, it is possible to build a finite state transducer which takes any input and returns the output which eval maps it to. It is a simple matter to convert this transducer into a finite state acceptor which accepts or rejects words. Thus, given a set of ranked OT constraints written as finite state machines, it is possible to compute a phonotactic grammar. This representation of the grammar implicitly embodies the constraints and the ranking. It is reasonable therefore to ask whether this acceptor can be learned directly from observed surface forms.

As an example, if the (different) OT analyses of the placement of primary stress in the Leftmost Heavy Otherwise Leftmost pattern given in Walker (2000) and Bakovic (2004) were encoded as finite state machines, Riggles algorithm would yield the (same) phonotactic acceptor for this pattern shown in Figure 1. Similarly, their analyses for the Leftmost Heavy Otherwise Rightmost pattern would produce the acceptor shown in Figure 2.

The learning algorithm here returns these two acceptors after exposure to a sample of words from these languages. In fact, for the 17 variations of primary stress placement in unbounded stress systems described in Baileys (1995) stress system database, the learner succeeds in every case.

The learner succeeds because of a previously unnoticed universal property of unbounded stress systems—they are neighborhood-distinct, just like quantity-insensitive (QI) stress patterns (Heinz 2006). Neighborhood-distinct languages, which form a finite subset of the regular languages, are defined to be those languages for which an acceptor exists such that each state has a unique neighborhood. The neighborhood of a state is defined as the set of incoming symbols, the set of outgoing symbols, and whether the state is final or not (Figure 3). The neighborhood is thus the immediate local environment of a state. Because unbounded stress systems are neighborhood-distinct, the learner that succeeds for QI patterns succeeds here. This is a surprising but welcome result since on the face of it unbounded stress systems appear more complex than QI ones.

The learner works in two stages. First, the learner builds a finite state representation of the input. This is a grammar which accepts only the observed words. The learner then generalizes to new forms by merging states which have the same neighborhood (cf. (Angluin 1982)). When two states are merged in a machine they are combined and transitions to and from each state are kept, which may result in generalization (see Figure 4). Because the learner merges all states with the same neighborhoods, the end result is a machine which is neighborhood-distinct.

Other approaches to locality in phonology use n-gram models which are not only complex but are inherently unable to capture the patterns exhibited in unbounded stress systems unless they operate on additional levels of representation, i.e. tiers. On the other hand, the grammars acquired here are quite simple, and it is striking that this formalization of locality allows the learner to

generalize correctly for every attested unbounded stress system.

Figures. The hexagon indicates the start state and double peripheries final states. L0 indicates a stressless light syllable, L1 a stressed light, H0 a stressless heavy, and H1 a stressed heavy.

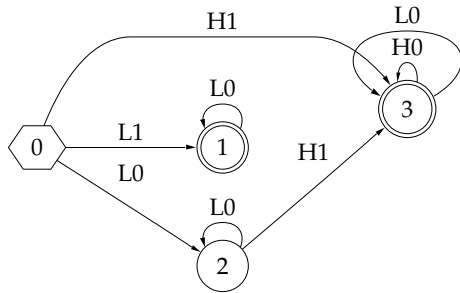


Figure 1: Leftmost Heavy Otherwise Leftmost

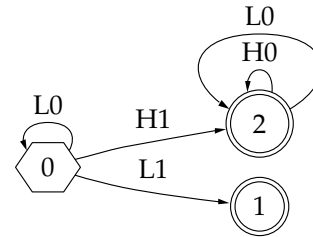


Figure 2: Leftmost Heavy Otherwise Rightmost

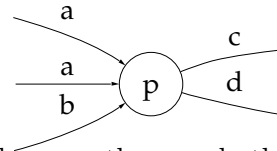
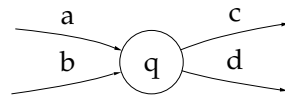


Figure 3: States p and q have the same neighborhoods because they are both nonfinal states, they can only be reached by symbols $\{a,b\}$ and they can only be departed from symbols $\{c,d\}$.



Figure 4: The right machine shows the result of merging states 1 and 2 in the left machine. The right machine preserves the transition from state 1 to 2 in the left machine resulting in generalization.

Selected References

- Albright, Adam and Bruce Hayes. 2003. Rules vs. Analogy in English Past Tenses: A Computational/Experimental Study. *Cognition* 90:119–161.
- Angluin, Dana. 1982. Inference of Reversible Languages. *Journal for the Association of Computing Machinery* 29(3):741–765.
- Bailey, Todd. 1995. Nonmetrical Constraints on Stress. Ph.D. thesis, University of Minnesota. Stress System Database available at <http://www.cf.ac.uk/psych/ssd/index.html>.
- Bakovic, Eric. 2004. Unbounded stress and factorial typology. In *Optimality Theory in Phonology: A Reader*, edited by John McCarthy. Blackwell, London. ROA-244.
- Hayes, Bruce. 1995. *Metrical Stress Theory*. Chicago University Press.
- Hayes, Bruce. 2004. Phonological acquisition in Optimality Theory: the early stages. In *Fixing Priorities: Constraints in Phonological Acquisition*, edited by Rene Kager, Joe Pater, and Wim Zonneveld. Cambridge University Press.
- Heinz, Jeffrey. 2006. Learning Quantity Insensitive Stress Systems via Local Inference. In *Proceedings of the ACL-SIGPHON* 8:21–30.
- Idsardi, William. 1992. The Computation of Prosody. Ph.D. thesis, MIT.
- Prince, Alan. 1983. Relating to the Grid. *Linguistic Inquiry* 14(1).
- Prince, Alan and Paul Smolensky. 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Blackwell Publishing.
- Riggle, Jason. 2004. Generation, Recognition, and Learning in Finite State Optimality Theory. Ph.D. thesis, University of California, Los Angeles.
- Walker, Rachel. 2000. Mongolian Stress, Licensing, and Factorial Typology. ROA-172.