Title
Learning OT Grammars of Syllable Structure

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 26(26)

ISSN
1069-7977

Author
Wayment, Adam T.

Publication Date
2004

Peer reviewed

Learning OT Grammars of Syllable Structure

Adam T. Wayment (awayment@jhu.edu)
Department of Cognitive Science, Johns Hopkins University
3400 N. Charles St. Baltimore, MD 21218 USA

1. Introduction

Optimality Theory (OT) (Prince & Smolensky 1993) has been widely adopted in phonology and has also been successfully applied to syntax, semantics, and pragmatics. One reason OT has been so rapidly accepted is that its initial presentation was closely tied to a connectionist realization (Blutner, et al. forthcoming). Goldwater and Johnson (2003) suggest that another reason for OT's recent dominance is that there are algorithms for learning constraint rankings. However, these elements of OT's success (learning algorithms and connectionist realizations) have not yet been unified in a connectionist network that learns constraint rankings.

A grammar in Optimality Theory is defined by a set of ranked violable constraints. The function GEN takes a base form as input and generates an infinite set of candidates. EVAL then selects the optimal realization from among these candidates, obeying the criterion that the ordering of constraints is strictly dominant. Archangeli (1997) reviews how re-ranking a few violable constraints (ONSET, PEAK, NOCODA, *COMPLEX, FAITHC, AND FAITHV) accounts for a large number of the syllable structures attested in the languages of the world. Within this domain of syllable structure, we explore the dilemma of learning a constraint ranking in a connectionist network. The goal is to design a network that can learn the well-formedness of test syllables, based on positive training data generated from a particular ranking of the violable constraints.

2. Fixed-Point Membership

Now, let **G** be a harmonic grammar. The task is to determine if an input form w is in $L(\mathbf{G})$. Define language membership as follows $w \in L(\mathbf{G})$ iff EVAL(GEN(w)) = w. In words, membership is equated with being a fixed-point of OT generation. This concept of membership provides a powerful framework, in which recognition can be performed via a fixed point test on an input form.

Translating this notion to a Harmonic Network—where input is equivalent to clamping the initial activation state—if, after the network is allowed to harmonize, the input pattern is the same as the output pattern, then the input form is in the language of the harmonic grammar described by the weights of the network. If however, the input and output pattern differ, then the input form is not in the language prescribed by the grammar because the more harmonic activation state corresponds to some other output form, so the form fails the fixed-point test. Thus, learning well-formedness is tantamount to learning the identity map. Poverty of the stimulus issues bear heavily on this problem because we require that the identity map be learned from only positive data.

3. Data, Representation, and Learning

Sample data consists of valid phonetic combinations of consonants and vowels, e.g. for a CV language, [ba], [mi], and [po] may be present in the training. Network evaluation is determined by way of the well-formedness scores of a random collection of test syllables of various structure (CV, CVC, CCV, etc.). Only those syllables that are in the language of the net's grammar will be fixed points, and all forms that are not in the language will not.

We represent syllables in a fully connected-symmetric network made by copying a set of fillers (one unit per consonant or vowel) for every role position (peak, onset, etc.). We allow for complex onsets and codas, by placing multiple filler sets in a position. Thus, 'tint' =[tInt], is represented by activating the *t* unit in the Onset filler set, the *I* unit in the Nucleus filler set, the *n* unit on in the first coda filler set, and the *t* unit in second coda filler set.

We compare two different algorithms for learning the weights of the network: backpropagation through time and Boltzmann learning. Preliminary experiments show that the first is unsuccessful at learning the elements in the complement of the training set are ungrammatical, whereas because of its negative phase of training, Boltzmann learning can learn the well-formedness of syllables.

4. References

Archangeli, D (1997) *Optimality Theory: An Overview*.

Blutner, R., Hendriks P., and de Hoop, H. (2003) *Optimal Forms and Meanings*. Book draft.

Goldwater, S. & Johnson, M. 2003. Learning OT Constraint Rankings Using a Maximum Entropy Model. In *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*. 111-20.

Prince, A. and Smolensky, P. 1993. Optimality theory: Constraint interaction in generative grammar. Technical Report TR-2, Rutgers Center for Cognitive Science, Rutgers Univ.