Distinguishing Grammars in Multilingual Learning Using Parameter Co-occurrence

Problem

If learners in multilingual environments are given samples from a mixture of grammars, how well can they distinguish between individual grammars using information contained in those samples? Optimality Theoretic accounts of free variation (Antilla 1997, Boersma & Hayes 2001) model scenarios where learners acquire *unions* of grammars. This differs from multilingual scenarios, which require learning *disjunctions* of grammars. For instance, in learning grammars G1 & G2, where G1 epenthesizes onsets and G2 deletes codas, the multilingual learner's challenge is to avoid generalizing to G3, which does both.

(1) a. G1: /VC/ → [CVC]
b. G2: /VC/ → [V]
c. *G3: /VC/ → [CV]

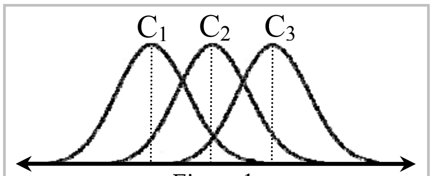

Figure 1

This is precisely the problem with using free variation models for multilingual learning. Fig. 1 shows a Boersma-Hayes style model of variation between constraint rankings $C_1 >> C_2 >> C_3$ and $C_3 >> C_2 >> C_1$. There is no way to prevent this model from generating all six rankings of $\{C_1, C_2, C_3\}$ (the probability of each permutation being determined by the overlap of the Gaussians).

Discrimination Algorithm

We propose a heuristic for distinguishing grammars in multilingual scenarios by partitioning observed grammatical parameters according to their co-occurrence. We assume a supervised learning scenario with input-output pairs drawn from multiple OT grammars, parameterized as sets of Elementary Ranking Conditions (ERCs; Prince 2002). Assuming universal CON, the learner can infer the set of ERCs consistent with a given I-O mapping. Our learning algorithm establishes a list of pairs of ERCs that co-occur in the ERC set of at least one I-O mapping in the training sample. These pairs define a network of *n* ERCs whose dense regions correspond to collections of parameters associated with each other in the samples (Fig. 2). We use these regions of strongly associated parameters as the basis for distinguishing grammars.

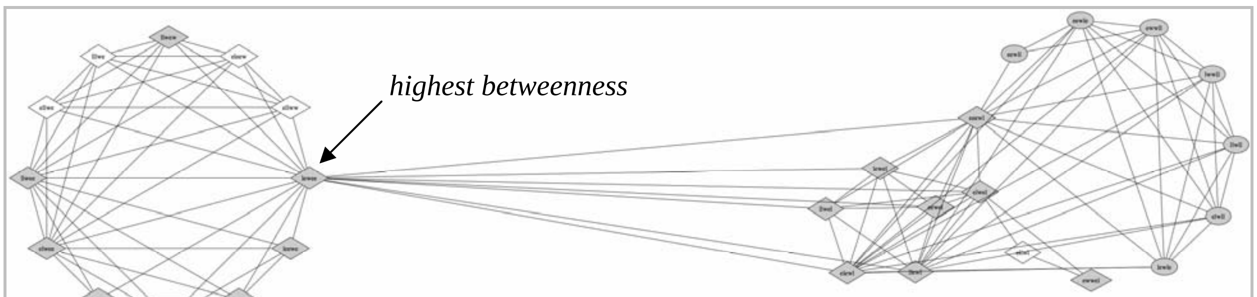

Figure 2: Three Dense Regions

The algorithm proceeds as follows: for each discrete ERC set, if that set is *consistent* (i.e. free of internal contradictions), it becomes one of the learner's hypotheses. Otherwise, the inconsistency is resolved by recursively removing ERCs with the highest *betweenness centrality* (Brandes 2001) in the network. The algorithm will converge on some number of discrete, internally consistent ERC sets after $O(n)$ iterations.

Results

We ran 5,000 trials with 1, 2, or 3 teachers using randomly generating 10-constraint syllable structure grammars. In each trial learners received from 10 to 265 training samples and generated *k* ERC sets (hypotheses) using our algorithm. These were rated in terms of average errors of the type in (1c). Fig. 3 shows the number of trials resulting in *k* hypotheses (bars) and average error rate across trials with *k* hypotheses (lines). While error increases with *k*, most trials generate relatively few hypotheses, suggesting the general utility of the heuristic.

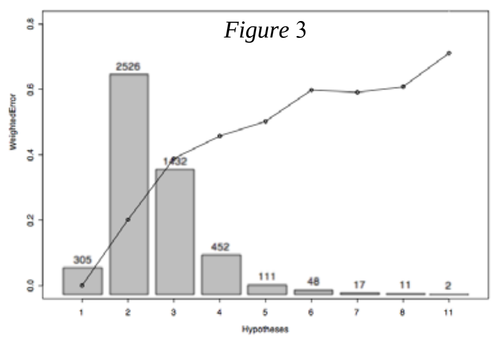

Figure 3