Chapter 5 from Sakas, W.G. (2000) Ambiguity and the Computational Feasibility of Syntax Acquisition. PhD Dissertation, City University of New York.

Discussion and Directions

The results presented in the dissertation provide indirect evidence that for two very different models, learning performance differs widely in domains with different distributions of ambiguity. Two factors emerge that exponentially affect the performance of syntax acquisition:

- the size of the parameter space, and
- the degree of parametric ambiguity[1]

For the TLA the size of the parameter space poses a major hurdle in the case that ambiguity is distributed evenly throughout the domain – the TLA performs notably worse than a learner that blindly guesses grammars at random, even at rates of ambiguity most favorable to TLA performance. The number of sentences required to converge on the target is hence greater than 2^n where n equals the number of parameters that need to be set. However, when the ambiguity is distributed so that a smooth domain ensues (i.e. the probability of a successful parse is not uniform, it is higher the closer a grammar is to the target grammar), the damaging effect of ambiguity is curtailed. At least as long as the number of parameters is modest, the TLA can acquire the target grammar consuming a reasonable number of input sentences.

Two potential concerns are not alleviated by this result. First, although the absolute number of sentences consumed by the TLA is reasonable, an exponential trend is exhibited. If the number of parameters that is required to accurately describe human language is greater than 30 (the largest domain tested in the dissertation), further research would be needed to determine if the number exceeds the point at which an exponential explosion occurs. It could be, for instance, that the TLA is a feasible model in smooth domains of up to 39 parameters, but a clearly infeasible one in a domain of 40. Second, the result (for strongly smooth domains with 30 or fewer parameters) is contingent on there being a language domain in which languages that are parametrically similar share an extremely large number of sentence types. That is, there needs to be a high degree of overlap between the languages generated by neighboring grammars in the grammar space. Again, if the TLA is to be considered a feasible model, future research would be required to determine if natural languages conform to the experimental boundaries within which efficient TLA performance was observed. In respect to language overlap, however, this seems unlikely.

For the waiting-STL, the Parametric Principle prevents the effect of the domain size from dominating performance. However, a high degree of parametric ambiguity is

[1] including the effect of expression rate on ambiguity.

crippling. That is, as the rate of ambiguity increases, the STL requires an exponentially increasing number of sentences in order to attain the target grammar. As the degree of cross-language ambiguity in natural language rises above a modest amount, the number of sentences consumed by the STL rapidly escalates at an unmanageable rate. However, as for the TLA, there is a mitigating circumstance in which the STL can be expected to consume a reasonable number of inputs. When the number of parameters that are expressed per sentence can vary across the sentences of the input sample so that there are some sentences with little or no ambiguity, there is a striking improvement in STL performance. So, even at extremely high rates of ambiguity the STL can be considered a feasible learning model.

Still, as with the TLA, there are several potential concerns. First, it may be that the degree of parametric ambiguity in human languages approaches 100%. That is, there may be very few (if any) sentences that belong to the target language and the target language alone. If this is indeed the case, the STL is effectively paralyzed at the outset; the first parameters would be nearly impossible to set. Second, although varying parametric expression rate across sentences allows the Parametric Principle to keep the effect of the domain size in check (relative to the effect of domain size on TLA performance), there is still an accelerating increase in the number of sentences that are consumed in domains where the expression rate is fixed and the ambiguity rate is high. If psycholinguistic research reveals that this is indeed the case – that in human language, the number of parameters expressed fluctuates by very little and the ambiguity rate is high – the study presented here would have to be tuned to match the psycholinguistic data in order to determine if the STL is feasible under these conditions.

A twofold picture can be constructed from the STL and TLA feasibility studies. First, given the wide breadth of factors that create hospitable learning environments for the different learners (e.g. smoothness for the TLA, varying expression for the STL) it seems very unlikely that any one parameter setting algorithm could be devised that performs well in all possible domains. In other words, the evidence suggests that there is no acquisition model that exhibits superior performance across a broad range of possible learning scenarios. I call this the ***No Best Strategy Conjecture: No learning strategy is generally the most effective for setting syntactic parameters across all ambiguous domains.***[2]

[2] This follows the spirit of Schaffer's Law of Conservation of Generalization (1994) which applies to classification learners. Roughly, the result indicates that identifying the correct category of a datum that the learner has not been trained on (by generalizing from other data) is a "zero-sum enterprise" — for every learner, positive performance in some class of learning situations is offset by equally poor performance in others. Of course, the No Best Strategy Conjecture is concerned with efficiency, not accuracy.

Second, the effect of ambiguity on learning performance is *extremely sensitive to the **way** in which the ambiguity emerges*. Any computational framework that is employed to simulate language acquisition requires that certain variables (e.g. α, λ, a, e, etc.) be in place that define the input sample presented to the learner. The precise formulation of these variables shapes the distribution of ambiguity, which in turn affects learning performance. Certain distributions prevent feasible learning while others allow it. Within the class of all possible learning situations, both the STL and the TLA have *sweet spots* - where the *shape of ambiguity* is favorable to learning performance. It is noteworthy that these situations are narrowly circumscribed; small changes in the relevant variables result in large changes in performance outcomes.

Together, this sensitivity and the No Best Strategy Conjecture strongly suggest that knowledge of successful performance in one artificial domain can not be used to argue for an algorithm's viability as a model of true natural language acquisition. Attempts at discovering the mechanism of human language acquisition through computational modeling must be coupled with a detailed analysis of the shape of ambiguity in input samples typically encountered by children. The results presented here contribute to a growing body of research indicating that parameter setting is a difficult and subtle enterprise. They also underline the fact that whether a particular acquisition model ultimately succeeds or fails will depend on the exact conditions under which the model performs well and the extent to which those favorable conditions are in line with the facts of human language.