ABSTRACT

Title of Document:	NECESSARY BIAS IN NATURAL LANGUAGE LEARNING
	Lisa Sue Pearl, Doctor of Philosophy, 2007
Directed By:	Associate Professor, Department of Linguistics/UMIACS Associate Director, Neural and Cognitive Science Program Co-Director, Laboratory for Language and Media Processing Amy Weinberg

This dissertation investigates the mechanism of language acquisition given the boundary conditions provided by linguistic representation and the time course of acquisition. Exploration of the mechanism is vital once we consider the complexity of the system to be learned and the non-transparent relationship between the observable data and the underlying system. It is not enough to restrict the potential systems the learner could acquire, which can be done by defining a finite set of parameters the learner must set. Even supposing that the system is defined by *n* binary parameters, we must *still* explain how the learner converges on the correct system(s) out of the possible 2^n systems, using data that is often highly ambiguous and exception-filled. The main discovery from the case studies presented here is that

learners can in fact succeed provided they are biased to only use a subset of the available input that is perceived as a cleaner representation of the underlying system.

The case studies are embedded in a framework that conceptualizes language learning as three separable components, assuming that learning is the process of selecting the best-fit option given the available data. These components are (1) a defined hypothesis space, (2) a definition of the data used for learning (data intake), and (3) an algorithm that updates the learner's belief in the available hypotheses, based on data intake. One benefit of this framework is that components can be investigated individually. Moreover, defining the learning components in this somewhat abstract manner allows us to apply the framework to a range of language learning problems and linguistics domains. In addition, we can combine discrete linguistic representations with probabilistic methods and so account for the gradualness and variation in learning that human children display.

The tool of exploration for these case studies is computational modeling, which proves itself very useful in addressing the feasibility, sufficiency, and necessity of data intake filtering since these questions would be very difficult to address with traditional experimental techniques. In addition, the results of computational modeling can generate predictions that can then be tested experimentally.