Evaluating constructivist theory via Bayesian modeling of children's early grammatical development

Theories of grammatical development differ in how much abstract knowledge they attribute to young children. While it has long been popular to discuss children's linguistic competence in terms of abstract categories and rules, a number of researchers have recently argued that in the absence of commitment to UG, it is inappropriate to pre-empirically assume continuity with adult language (e.g. Goldberg, 2006; Tomasello, 2003). Such usage-based researchers have proposed that children's speech for at least the first 2 years of multiword speech is remarkably restricted, with constructions being seen with only a small set of frequent verbs (Tomasello, 1992) and many utterances being built from lexically-specific frames (e.g. Pine and Lieven, 1993; Lieven et al, 2003). We here report on a series of experiments in which we use a computational model to evaluate the explanatory power of such a model of early production.

We recorded 2 English-speaking children for an hour on each of 5 days a week for 6 weeks immediately following their second and third birthdays: 30 hours of recordings for each child at each age. We employed a Bayesian procedure to extract grammars from the child's (not caregiver's) speech in the first 28 hours of each of these transcriptions, setting the remainder aside. Our basic grammars were formally equivalent to context-free grammars (Chomsky, 1956), recognizing the hierarchical nature of language and yet they consisted entirely of lexically-specific constructions - schematic patterns, words and multiword sequences.

In a first experiment we used these grammars and a standard algorithm (Kasami, 1965) to parse all the child's unique multiword utterances from transcriptions of the remaining two hours of speech for each child at each age. We found that at 2 years the children's productions could be accounted for effectively (as much as 87% coverage) and perspicuously (as much as 65% of utterances requiring application of two rewrite rules or less) using such a concrete model. In a second experiment we examined the explanatory value of positing that the children had knowledge of the abstract categories of noun and of verb. We found that at age 2 adding a noun category produced a significant improvement in performance but adding a verb category had almost no impact. At age 3 the addition of abstract linguistic information over verbs as well as nouns markedly improved performance. These results are consistent with experimental findings regarding children's early knowledge of linguistic categories (Olguin and Tomasello, 1993). Taken as a whole our results support an account of early language development in which abstract linguistic knowledge develops gradually.