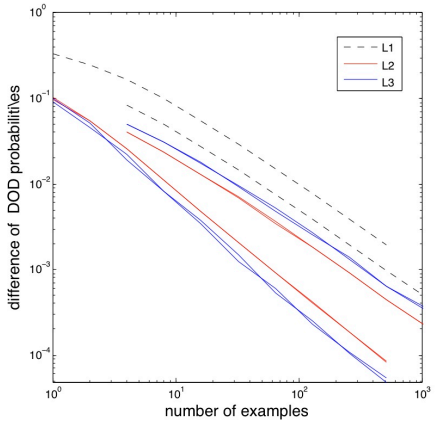
Keep it Simple: Language Acquisition Without Complex Bayesian Models

The work of Hsu & Chater [1] and Perfors et al. [2] establishes that sophisticated statistical learning techniques known as Hierarchical Bayesian Models (HBMs) can successfully capture certain observed patterns of both under- and over-generalization in child language acquisition. This paper shows that a much simpler method, maximum likelihood estimation (MLE), can equal HBM performance. The work in [2] analyzed dative alternations compiled from child-directed CHILDES English or from controlled language experiments (Wonnacott et al., [3]). However, HBMs are 'ideal' learning systems, known to be computationally infeasible (Kwisthout et al. [4]). Consequently, as [4] notes, the relevance of HBMs for cognitively plausible accounts of human learning remains uncertain. This paper that combining simple clustering methods along with MLE provides an alternative, more cognitively plausible account of the same facts.

It has long been recognized that children manifest subtle patterns of under- and over-generalization with respect to learning dative verb alternation frames, using a combination of both verb-particular information as well as general verb-class behavior (Baker, 1979 [5]; Groppen et al, 1991 [6]), e.g., *John told the police the story/told the story to the police*, but **confessed the police the story*. The HBM approach of [2] posits three levels of statistical estimation to capture this observed behavior, from counts of individual verb occurrence in direct object dative frames (DOD), prepositional dative frames (PPD), or alternating (both); to the frequency of frames themselves; to, finally, the *hierarchical* estimate of whether the alternation frames themselves are distributed uniformly or not. Observed counts in the Childes corpus may then used to estimate whether an unseen verb will be DOD, PPD, or alternating. Is such complexity needed? We re-analyzed the child-directed counts of the frames for 19 verbs (*give, say, ..., mail*) taken from the CHILDES Adam corpus as in [2], as well as subcat frame counts for these 19 verbs from all of the English CHILDES, approx. 32,000 examples altogether. We tested a total of 18 different non-HBM models, using several clustering methods (the latter implemented in the Weka package [7]). *K*-means clustering easily placed the verbs into one of 3 groups, while a smoothed maximum likelihood estimate (MLE) using these groups yielded dative frame predictions closely matching the performance of HBMs. Fig. 1 illustrates. Dotted lines show the simplest model's performance, while solid lines are HBM variants. The y-axis plots the log deviation between MLE and HBM estimates while the x-axis plots # of example instances.



Children may well be capable of powerful statistical reasoning, but our results and parsimony suggest that computationally simpler statistical abilities should first be ruled out before resorting to computationally infeasible, and cognitively less defensible, approaches. **(438 words)**

References

[1] Hsu, A.S. and N.Chater. 2010. The logical problem of language acquisition: A probabilistic perspective. Cognitive Science, 34:6, 972-1016.

[2] Perfors, A., J.B. Tenenbaum, and E. Wonnacott. 2010. Variability, negative evidence, and the acquisition of verb argument constructions. Journal of Child Language, 37, 607-642.

[3] Wonnacott, E., E.L. Newport, and M.K. Tanenhaus. 2008. Acquiring and processing verb argument structure: Distributional learning in a miniature language. Cognitive Psychology, 56, 165-209.

[4] Kwisthout, J. Wareham, T., Rooijc, I. 2011. Bayesian intractability is not an ailment that approximation can cure. Cognitive Science 35, 779–784.

[5] Baker, C. (1979). Syntactic theory and the projection problem. Linguistic Inquiry, 10, 533–581.

[6] Gropen, J., Pinker, S., Hollander, M. & Goldberg, R. (1991). Syntax and semantics in the acquisition of locative verbs. Journal of Child Language 18:1, 115–151.

[7] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I. 2009. The WEKA data mining software: an update. SIGKDD Explorations, 11:1.