Learning English Metrical Phonology: Beyond Simple Probability

Language learning is a notoriously tricky business. There are multiple kinds of knowledge a child must learn, some more transparently related to the observable linguistic data than others. An example of the latter is a complex linguistic system; the child must hypothesize the correct system *components* (e.g. movement rules in syntax) while only ever encountering examples of the system's *output* (the observable data). Recent computational modeling work (syntax: Pearl & Weinberg, 2007; metrical phonology: Pearl, 2008) has suggested that children can succeed in some cases if they have an interpretive bias for the available data that causes them to selectively learn from only a portion of it. Yet is such a bias really necessary, given the strength and flexibility of probabilistic learning? Here I will demonstrate that several psychologically plausible probabilistic learners without interpretive biases fail to learn the parametric metrical phonology system studied previously by Pearl (2008), given realistic English data. This contrasts strongly with probabilistic learners using an interpretive bias, who are guaranteed to succeed given this same data. These results highlight the necessity for something beyond simple probabilistic learning when inferring a complex linguistic system.

Because human learning is gradual and reasonably robust to noise in the data, some kind of probabilistic learning is likely necessary. However, without constraints on what linguistic systems are possible, the hypothesis space for the learner is *infinite* – and even a sophisticated probabilistic learner will have difficulty choosing the correct system. Linguistic theory provides one idea for how to constrain the learner's hypothesis space: learners are guided by a selection of parameters (Chomsky,1981; Halle & Vergnaud, 1987) or constraints (Tesar and Smolensky, 2000) that enumerate the range of possible systems in human languages. The child's task then is to converge on the correct linguistic system within this subset, using the native language data. But the task is still not easy – the available data are often ambiguous and exception-filled. Learners who have an interpretive bias may have an advantage; the bias focuses their attention on highly informative data (Fodor, 1998; Lightfoot, 1999; Dresher, 1999).

The current modeling study explores the performance of probabilistic learners who have constraints only on the hypothesis space of possible systems, but no interpretive biases for the data. Each of the learning algorithms examined is incremental (adapted from Yang (2002)), predicated on children integrating information about their language as they encounter it. The complex system presented is a parametric instantiation of the English metrical phonology system (adapated from Hayes (1995) and Dresher (1999)), and involves 9 interacting parameters. The input data set is extrapolated from English child-directed speech (CHILDES: MacWhinney 2000), and is highly noisy. Perhaps surprisingly, it turns out that a probabilistic learner without an interpretive bias cannot succeed reliably, even when the hypothesis space is tightly constrained. Something additional is required, whether in the form of even tighter constraints on the hypothesis space or an interpretive bias for the data. In any case, simple probabilistic learning alone will not generate children's successful learning behavior.

Word count: 499

REFERENCES:

Chomsky, N. (1981). *Lectures on Government and Binding.* Dordrecht: Foris.

Dresher, E. (1999). Charting the learning path: Cues to parameter setting. *Linguistic Inquiry*, *30*, 27-67.

Fodor, J. D. (1998). Unambiguous Triggers. *Linguistic Inquiry, 29*, 1-36.

Halle, M. & Vergnaud, J-R. (1987). *An essay on stress.* Cambridge, MA: The MIT Press.

Hayes, B. (1995). *Metrical Stress Theory: Principles and Case Studies.* Chicago: University of Chicago Press.

Lightfoot, D. (1999). *The Development of Language: Acquisition, Change, and Evolution.* Oxford: Blackwell.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk.* Mahwah, NJ: Lawrence Erlbaum Associates.

Pearl, L. (2008). Putting the Emphasis on Unambiguous: The Feasibility of Data Filtering for Learning English Metrical Phonology, *BUCLD 32: Proceedings of the 32nd annual Boston University Conference on Child Language Development,* Cascadilla Press.

Pearl, L., & Weinberg, A. (2007). Input Filtering in Syntactic Acquisition: Answers from Language Change Modeling. *Language Learning and Development, 3(1),* 43-72.

Tesar, B. & Smolensky, P. (2000). *Learnability in Optimality Theory.* Cambridge, MA: The MIT Press.

Yang, C. (2002). *Knowledge and Learning in Natural Language.* Oxford: Oxford University Press.