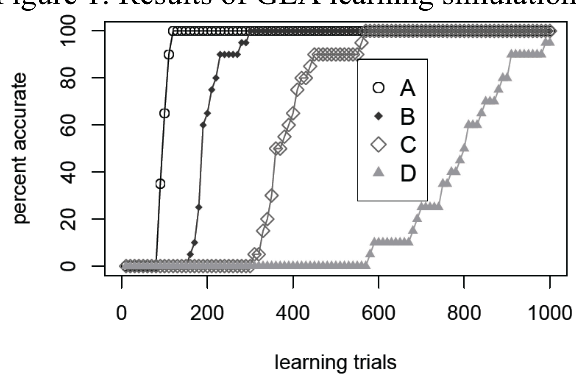
On the Relationship between Learning Sequence and Rate of Acquisition

Karen Jesney • University of Southern California

Overview: The Gradual Learning Algorithm (GLA; Boersma 1998, Boersma & Hayes 2001) predicts that more frequent input forms will be acquired earlier than less frequent input forms – a fact that has been commonly taken as a virtue of the model (e.g., Boersma & Levelt 2003, Curtin & Zuraw 2002, Jarosz 2010). The GLA also predicts, however, that the *rate* of learning for more frequent input forms should be *faster* than the rate of learning for less frequent input forms. In other words, the model predicts that sequence and rate of acquisition are related; structures acquired earlier in the course of learning will be acquired more rapidly, while those that are acquired relatively later will be acquired more slowly. This paper explicates these predictions and argues that they are not consistently supported by child language data.

Predictions of the learning model: The relationship between sequence and rate of acquisition in the GLA stems from two key properties of the model: the learner is error driven, and, in common implementations, target forms are sampled in accordance with their probability in the target language. As a result, the ranking values of constraints associated with more frequent target forms are adjusted more often than are those associated with less frequent target forms. This means that more frequent forms are mastered earlier *and* that the progression from the first accurate realizations to 100% accuracy occurs more rapidly for more frequent forms.

Figure 1: Results of GLA learning simulations



To illustrate this effect I constructed a toy language with the four forms /A/, /B/, /C/ and /D/, each of which provides evidence about the ranking of a markedness constraint (*A, *B, etc.) and a conflicting faithfulness constraint (FAITH-A, FAITH-B, etc.). An initial M >> F ranking was assumed. Figure 1 shows the mean results of 10 GLA simulations conducted in Praat (Boersma & Weenink 2014) based on an input distribution where the probability of /A/ was twice the probability of /B/, the probability of /B/ was twice the probability of /C/, etc. Mappings for each input form were sampled after every ten pieces of learning data. As expected, the most frequent input form, /A/, began to be realized accurately first, while the least frequent input form, /D/, was the last to begin to be realized accurately. Furthermore, it took an average of only 40 learning trials for input /A/ to shift from less than 10% accurate realization to over 90% accurate realization, while it took 320 learning trials for input /D/ to make the same transition. Inputs /B/ and /C/ fell between inputs /A/ and /D/ in terms of both sequence and rate of learning.

Child language data: Longitudinal corpus data allows us to test whether the predicted relationship between sequence and rate of acquisition holds consistently in child language. Data from two English-acquiring children are considered here: Trevor (Compton & Streeter 1977, Pater 1997) and Amahl (Smith 1973). For each child all target utterance-initial onset clusters and utterance-final coda clusters were extracted from the corpus. Target clusters found in unstressed syllables and those formed through morphological concatenation were excluded. For Trevor this yielded a total of 1633 tokens distributed across 40 cluster types (rhotic dialect), while for Amahl it yielded a total of 1496 tokens distributed across 51 cluster types (non-rhotic dialect). Target clusters were coded as accurate if they were produced as a sequence of two consonants, regardless of segmental changes. Additional details are given in the table on page 2.

	Trevor (0;11-3;1)		Amahl (2;2-3;9)	
	onset	coda	onset	coda
rising sonority	877	63	845	44
falling sonority	120	573	199	408
total	997	636	1044	452

For both Trevor and Amahl, the probability of all clusters being realized accurately increased significantly with age. In both cases, however, a logistic regression model with the factors age, syllable position, and sonority fully crossed provided a better fit to the data than any simpler model (Trevor: $p < .01$, Amahl: $p < .001$). This indicates that the *rate* of acquisition varied across cluster position and sonority profile.

Figures 2 and 3 plot the predicted probability of accurate realization for the different cluster types based on the fitted logistic regression models. For Trevor, the overall pattern largely mirrors that predicted by the GLA. The cluster types that Trevor begins to produce earliest – rising and falling sonority coda clusters – reach a high level of accuracy at a faster rate than the later-acquired cluster types. For Amahl, on the other hand, the pattern directly contradicts the predictions of the GLA. As Figure 3 shows, falling sonority onset clusters are the last cluster type that Amahl begins to realize accurately, but his rate of acquisition for this cluster type is more rapid than for any other type.

Figure 2: Trevor's accurate cluster realization

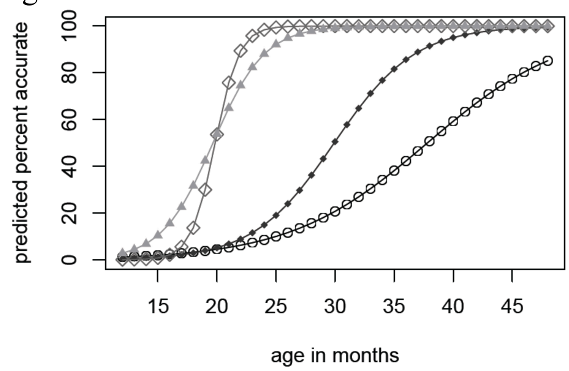
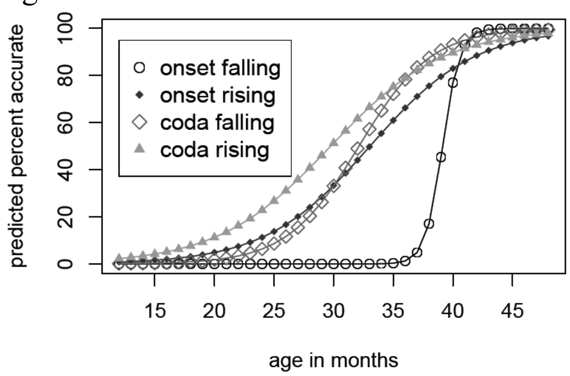


Figure 3: Amahl's accurate cluster realization



Implications: The predictions of the GLA outlined here extend to all gradual error-driven learning models that sample based on frequency (e.g., Noisy Harmonic Grammar – Boersma & Pater to appear, MaxEnt-OT – Goldwater & Johnson 2003). Comparisons with child data, however, indicate that the relationship between sequence and rate of acquisition is not as straightforward as these models predict. This points to the necessity of incorporating other factors, such as input restructuring and lexical growth, into models of phonological learning.

References: Boersma, P. 1998. *Functional Phonology: Formalizing the Interaction Between Articulatory and Perceptual Drives*. The Hague: Holland Academic Graphics. • Boersma, P. & B. Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32: 45-86. • Boersma, P. & C. Levelt. 2003. Optimality Theory and phonological acquisition. *Annual Review of Language Acquisition* 3: 1-50. • Boersma, P. & J. Pater. to appear. Convergence properties of a gradual learning algorithm for Harmonic Grammar. In J. McCarthy & J. Pater (eds.), *Harmonic Grammar and Harmonic Serialism*. London: Equinox Press. • Boersma, P. & D. Weenink. 2014. *PRAAT: Doing Phonetics by Computer*. [www.praat.org]. • Compton, A.J. & M. Streeter. 1977. Child phonology: data collection and preliminary analyses. In *Papers and Reports in Child Language Development* 13. Stanford University. • Curtin, S. & K. Zuraw. 2002. Explaining constraint demotion in a developing system. In A. Do, L. Domínguez & A. Johansen (eds.), *Proceedings of BUCLD 26*. Somerville, MA: Cascadilla Press. • Goldwater, S. & M. Johnson. 2003. Learning OT constraint rankings using a Maximum Entropy model. In *Proceedings of the Workshop on Variation within Optimality Theory*, 111-120. Stockholm University. • Jarosz, G. 2010. Implicational markedness and frequency in constraint-based computational models of phonological learning. *Journal of Child Language* 37: 565-606. • Pater, J. 1997. Minimal violation in phonological development. *Language Acquisition* 6(3): 201-253. • Smith, N.V. 1973. *The Acquisition of Phonology: A Case Study*. Cambridge: CUP.