

# On the Relationship Between Learning Sequence and Rate of Acquisition

Karen Jesney

*University of Southern California*

## 1 Introduction

The relative frequency of forms in child-directed speech has been claimed to play an important role in determining the sequence in which these forms are acquired. Levelt, Schiller & Levelt (1999/2000), for example, demonstrate that simple codas are acquired earlier than onsetless syllables by Dutch children, and argue that this is due to the significantly higher frequency of codas than of onsetless syllables in child-directed Dutch. These same children show no consistent sequence of acquisition for onset clusters vs. coda clusters – two structures that are of comparable frequency in the input data. Similar findings are noted by many others, including Roark & Demuth (2000), who observe that coda consonants are acquired earlier than weak syllables by English-learning children, consistent with the relative frequency of these structures in child-directed English, but that the opposite sequence of acquisition is observed among Spanish-learning children, where the opposite input frequency patterns hold. Additional arguments for the role of frequency in determining sequence of acquisition come from, among many others, Beckman & Edwards (2010), Edwards & Beckman (2008), Edwards, Beckman & Munson (2015), Demuth & Johnson (2003), Ingram (1989), and Jarosz (2010).

Less clear is the relationship between input frequency and rate of acquisition – i.e., the slope of the learning curve. The child language literature has generally recognized that not all structures show the same type of progression. Menn (2004), for example, distinguishes between instances of ‘across-the-board’ changes, where a marked structure goes from 0% accuracy to 100% in an essentially categorical manner, and ‘gradual’ changes, where the transition from 0% to 100% accuracy is much less abrupt. To my knowledge, however, the relationship of these patterns to input frequency has not been systematically explored.

Sequence of acquisition and rate of learning are of particular interest in the context of constraint-based error-driven learning algorithms. A direct, positive correlation between input frequency and learning sequence is predicted by a range of models, including the Gradual Learning Algorithm for Optimality Theory (OT-GLA; Boersma 1998, Boersma & Hayes 2001) and Noisy Harmonic Grammar (HG-GLA; Boersma & Pater to appear, Jesney & Tessier 2011), and Maximum Entropy OT (MaxEnt-OT) learned as in Goldwater & Johnson (2003). Indeed, the ability of these models to capture the relationship between frequency and sequence of acquisition has often been taken as support for this basic approach to phonological learning (e.g., Boersma & Levelt 2000, 2003, Curtin & Zuraw 2002). Importantly, these same learning models also predict that higher input frequency should also be associated with a more rapid rate of acquisition – i.e., a sharper learning curve. Taken together, the result is the set of predictions laid out in (1).

- (1) a. More frequent structures should begin to be realized accurately earlier than less frequent structures.
- b. Once initial accurate productions emerge, more frequent structures should progress toward full accuracy at a faster rate than less frequent structures.

---

\* Thank you to the audiences at USC and at AMP 2015 for their helpful comments and feedback on this work. Thank you also to all of the researchers who have made their corpora of child phonological data available through CHILDES, allowing projects of this type to be undertaken.

- c. **Rate-Sequence Prediction:** Sequence of acquisition and rate of learning should be correlated, such that structures which emerge earlier in acquisition should progress toward full accuracy at a faster rate than structures which emerge later in acquisition.

This paper assesses the support for these claims by looking at the longitudinal acquisition of onset and coda clusters with different sonority profiles. Overall, the results suggest that the relationship between learning rate and sequence of acquisition is tenuous at best. In many cases, later-acquired structures actually show more rapid development than earlier-acquired structures – a finding directly contrary to the Rate-Sequence Prediction of error-driven gradual learning models. This suggests that the close association of rate and sequence to input frequency in these models is unmerited, and that a more nuanced understanding of the learning process is desirable.

The rest of this paper is structured as follows. Section 2 presents the logic behind the predictions in (1), and illustrates the expected patterns using toy language data. Section 3 presents data from six English-acquiring children to test these claims, and finds that they are largely unsupported. Section 4 reconsiders the role of frequency in modeling these data, and section 5 concludes.

## 2 Model predictions

The gradual error-driven models of learning that are of interest here operate according to the basic procedure in (2).

- (2) Basic error-driven learning procedure
- a. Forms are sampled from a distribution that reflects their frequency in the target language.
  - b. The learner determines whether her current grammar will accurately reproduce the sampled form.
    - i. If yes, the grammar is unchanged.
    - ii. If no, the grammar is adjusted by:
      - increasing the value of constraints favouring the target form, and
      - decreasing the value of constraints favouring the current optimum.
  - c. The learner returns to step a. and continues until her grammar consistently reproduces the target forms.

This procedure is typically supplemented by the assumption that Markedness constraints are assigned a high initial value and Faithfulness constraints are assigned a low initial value. This favours the learning of restrictive final grammars, and also allows intermediate stages of child language acquisition to be more effectively modeled (Demuth 1995, Gnanadesikan 2004, Hayes 2004, Pater 1997, Smolensky 1996, Tessier 2007).

The primary determinant of both sequence and rate of acquisition in this model is the distribution of inputs forms to which the learner is exposed. More frequent forms are more likely to be sampled by the learner, and the constraints responsible for their accurate mapping are therefore subject to relatively more frequent adjustments to their values. Frequent adjustments lead to both earlier accurate production and faster progression from 0% to 100% accuracy.

To illustrate this effect, we can consider a toy language like that summarized in (3). There are four input forms in this language. The most frequent input /A/ is twice as frequent as the next most frequent input /B/; /B/ is twice as frequent as /C/; and /C/ is twice as frequent as /D/. A unique Markedness constraint militates against each input, and a unique Faithfulness constraint favours the faithful mapping of each input. Given the learning procedure in (2) and an initial Markedness over Faithfulness bias, errors involving any of these inputs will trigger a decrease in the value of the associated Markedness constraint and an increase in the value of the associated Faithfulness constraint.

## (3) Toy Language 1

| Input | Probability | Effect of error |                      |
|-------|-------------|-----------------|----------------------|
|       |             | Markedness      | Faithfulness         |
| /A/   | .533        | ↓ $w(*A)$       | ↑ $w(\text{FAITHA})$ |
| /B/   | .267        | ↓ $w(*B)$       | ↑ $w(\text{FAITHB})$ |
| /C/   | .133        | ↓ $w(*C)$       | ↑ $w(\text{FAITHC})$ |
| /D/   | .067        | ↓ $w(*D)$       | ↑ $w(\text{FAITHD})$ |

The logic yielding the Rate-Sequence Prediction is straightforward. Because input /A/ is sampled more frequently than any other form in the language, it is subject to more frequent errors, and the values of \*A and FAITHA change relatively rapidly. Accurate productions of input /A/ thus become possible fairly soon after learning begins. As input /A/ continues to be sampled, rapid learning continues and consistent accuracy on input /A/ soon emerges. The time lag between the initial accurate productions and full accuracy is short; learning appears to occur in an ‘across-the-board’ fashion. The opposite is true for the infrequent input form /D/. Given that errors implicating \*D and FAITHD are relatively rare, it takes longer for accuracy on input /D/ to become possible at all. Then, the period of variable realization for /D/ is relatively long lasting, as evidence for assigning FAITHD a value well above that of \*D only gradually accrues.

This predicted relationship between sequence and rate of acquisition extends directly to cases where more complex constraint violation profiles are at play. As illustration, we can consider the toy language sketched in (4). In (4) the violations assigned by the Markedness constraints intersect, so that \*AORB penalizes both A and B, \*BORD penalizes both B and D, and \*CORD penalizes both C and D. Because of the constraint violations overlap, multiple inputs can provide evidence about the necessary constraint values. In (4), for instance, fully 80% of the input forms indicate that the value of \*AORB should be decreased. The overall result is that learning is accelerated relative to the toy language in (3); forms are acquired earlier and the rate of learning is increased. The expected correlation between sequence and rate of acquisition does not change.

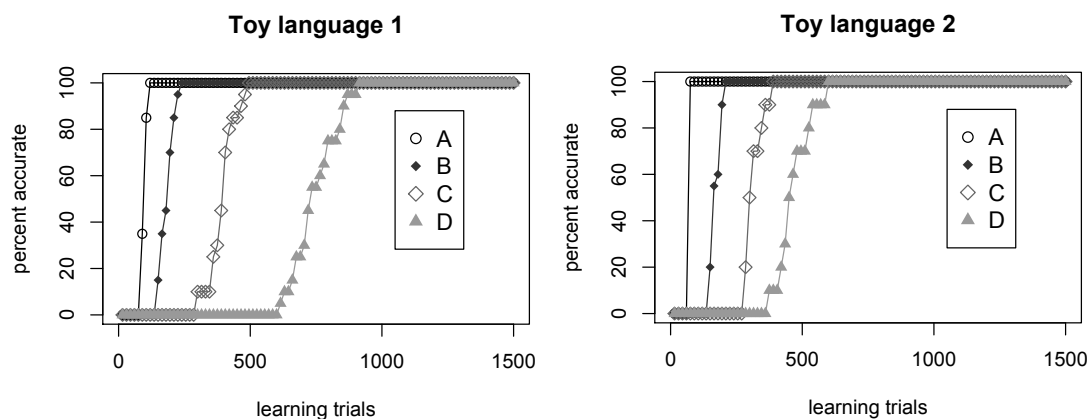
## (4) Toy language 2

| Input | Probability | Effect of error             |                      |
|-------|-------------|-----------------------------|----------------------|
|       |             | Markedness                  | Faithfulness         |
| /A/   | .533        | ↓ $w(*AORB)$                | ↑ $w(\text{FAITHA})$ |
| /B/   | .267        | ↓ $w(*AORB)$ , ↓ $w(*BORD)$ | ↑ $w(\text{FAITHB})$ |
| /C/   | .133        | ↓ $w(*CORD)$                | ↑ $w(\text{FAITHC})$ |
| /D/   | .067        | ↓ $w(*CORD)$ , ↓ $w(*BORD)$ | ↑ $w(\text{FAITHD})$ |

The basic relationship between sequence and rate of acquisition is illustrated in Figure 1 with MaxEnt-OT learning simulations for the two toy languages. Each figure shows the mean result of ten MaxEnt-OT simulations conducted in Praat (Boersma & Weenink 2012), with the predicted output distribution sampled after every 15 pieces of learning data.<sup>1</sup> The pattern is highly consistent for both toy languages. In each case, across the 10 separate learning simulations, the most frequent structure /A/ is consistently acquired first, followed by /B/, then /C/, then /D/. Furthermore, the earlier acquired structures consistently take fewer learning trials to progress to 100% accuracy once the initial accurate mappings occur. The two toy languages differ in the overall rate of acquisition – structures are acquired more quickly in the case of overlapping violations – but the relationship between sequence and rate of acquisition remains consistent.

<sup>1</sup> All Markedness constraints were assigned an initial weight of 100 and all Faithfulness constraints were assigned an initial weight of 0 (for this implementation of the M > F bias in weighted constraint systems, see Jesney & Tessier 2011). Additional simulation parameters: Decision strategy = “ExponentialMaximumEntropy”, Evaluation noise = 0.0, Update rule = Symmetric all, Initial plasticity = 1.0, Number of plasticities = 1, Number of chews = 1.

Figure 1: Mean results of MaxEnt-OT simulations showing the learning path for two toy languages



### 3 Child language data

**3.1 Overall relationship between sequence and rate** Assuming that gradual error-driven learning models accurately reflect the role of input frequency in the phonological learning process, patterns like those in Figure 1 should be widely attested in longitudinal studies of child language acquisition. This paper tests this claim by examining data from six English-acquiring children whose corpora are available through CHILDES (MacWhinney 2000). Details about each of the children are provided in (5).

(5)

| Child  | Age range   | Type of data               | Source                              |
|--------|-------------|----------------------------|-------------------------------------|
| Amahl  | 2;2 – 3;9   | diary study                | Smith 1973                          |
| Trevor | 0;11 – 3;1  | diary study                | Compton & Streeter 1977, Pater 1997 |
| E.     | 1;0 – 3;9   | diary study                | Inkelas & Rose 2003, 2008           |
| Alex   | 1;5 – 3;7   | regular recording sessions | Demuth, Culbertson & Alter 2006     |
| Ethan  | 0;11 – 2;11 | regular recording sessions | Demuth, Culbertson & Alter 2006     |
| Lily   | 1;2 – 4;0   | regular recording sessions | Demuth, Culbertson & Alter 2006     |

For the purposes of this study, four target structures were identified: rising sonority onset clusters, falling sonority onset clusters, rising sonority coda clusters, and falling sonority coda clusters. Only target clusters in stressed syllables were considered. To further control for syllabification, onset clusters were restricted to those in utterance-initial position and coda clusters were restricted to those in utterance-final position. Coda clusters formed through suffixation were excluded. The total number of attempts at each of these structures by each child is given in (6).

(6)

| Child        | Target onset clusters |                  | Target coda clusters |                  |
|--------------|-----------------------|------------------|----------------------|------------------|
|              | Rising sonority       | Falling sonority | Rising sonority      | Falling sonority |
| Amahl        | 845                   | 199              | 44                   | 408              |
| Trevor       | 877                   | 120              | 63                   | 573              |
| E.           | 186                   | 50               | 20                   | 123              |
| Alex         | 1177                  | 109              | 216                  | 561              |
| Ethan        | 742                   | 200              | 55                   | 581              |
| Lily         | 923                   | 152              | 128                  | 1234             |
| <b>Total</b> | <b>4750</b>           | <b>830</b>       | <b>526</b>           | <b>3480</b>      |

The cluster types attempted by the children in the four categories are given in (7); not all types were attempted by all of the children.

|     |   |  |
|-----|---|--|
| (7) | Rising sonority onset clusters:               | pl, pr, tr, tw, tj, kl, kr, kw, kj, bl, br, bw, bj, dr, gl, gr, fl, fr, fj, θr, θw, sl, sm, sn, sw, fr, fm, hj, vr, mw, nj                           |
|     | Falling sonority onset clusters:              | sp, st, sk   |
|     | Rising sonority coda clusters:                | ps, ts, ks, bz, dz, gz   |
|     | Falling sonority coda clusters <sup>2</sup> : | lp, lb, lt, ld, lk, lf, lv, ls, lm, rp, rb, rt, rd, rk, rtf, rɕ, rf, rθ, rs, rz, rm, rn, rl, mp, mθ, mz, nt, nd, ntʃ, nɕ, ns, nʒ, ŋk, ft, sp, st, sk |

While frequency mediates the expected relationship between sequence and rate of acquisition, it is unclear what type of frequency data should be considered most important. There are number of possibilities: the frequency of structures in child-directed speech, the frequency of attempts at structures by the learner, etc. We briefly return to these issues in section 4. Regardless, gradual error-driven learning models predict that there should be a clear association between the sequence and rate of acquisition. Structures whose accurate production becomes possible earlier should move toward full accuracy at a faster rate.

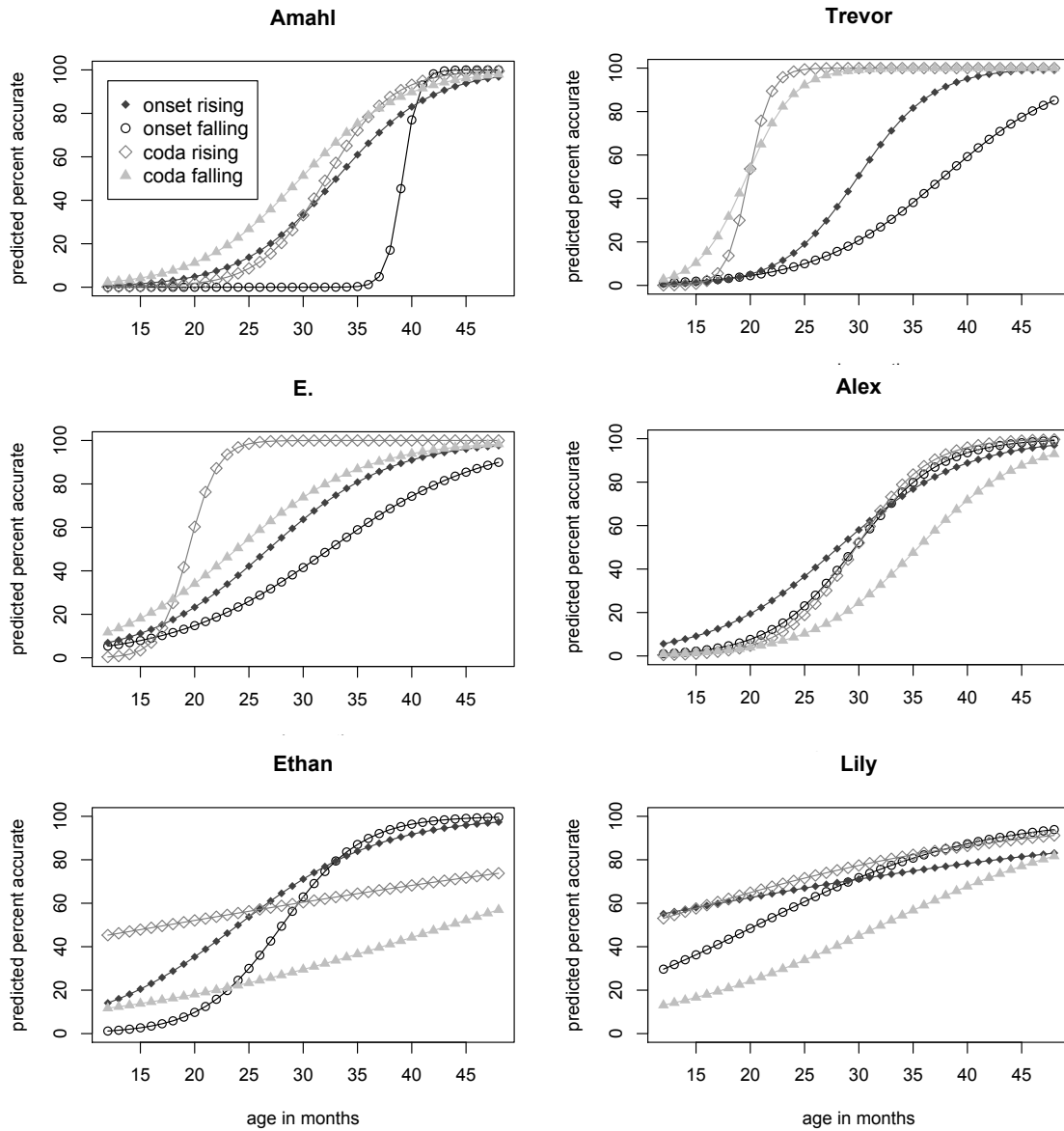
To test this association, all of the children's attempts at target onset and coda clusters were coded for accuracy. Productions were considered to be accurate if two consonants appeared in sequence without deletion or epenthesis. Tokens with errors only in segmental features (e.g., devoicing or changes in place of articulation) were coded as accurate. Separate logistic regression models were then fitted to each child's data using the *glm* function in R (R Core Team 2014). The factors age (in months), syllable position (onset vs. coda), and sonority profile (good vs. bad given the syllable position) were all found to be predictors of cluster accuracy. Fully-crossed models provided a consistently better fit to the data than less complex models (all  $p < .01$ ). Figure 2 plots the predicted percent accuracy for each of the cluster types for each of the children based on the best-fit models. Full statistical models are provided in the Appendix.

As is evident from these plots, there is little consistent relationship between sequence and rate of acquisition in these children's data. Overall, Trevor's onset data is arguably most consistent with the Rate-Sequence Prediction of the error-driven learning models, as later acquired structures (falling sonority onset clusters) show a slower rate of progression toward accuracy than earlier acquired structures (rising sonority onset clusters). Even for Trevor, though, the two coda cluster types invert the expected pattern; falling sonority coda clusters are acquired earlier than falling sonority coda clusters, but rising sonority coda clusters show a more rapid rate of acquisition. Other children's data also show inversions of the predicted patterns. Ethan, for example, shows relatively high levels of accuracy for rising sonority coda clusters from early on, but these never reach a predicted 100% accuracy rate within the time frame considered here. Falling sonority onset clusters, on the other hand, appear later but progress toward full accuracy at a faster rate than any of the other structures.

---

<sup>2</sup> Coda clusters including /r/ were excluded in the case of Amahl (Smith 1973), who was acquiring a non-rhotic dialect of English.

Figure 2: Predicted accuracy rate of cluster types by age in months for six English-acquiring children



In sum, at least at this level of granularity, there is little support for the Rate-Sequence Prediction of error-driven gradual learning models. This lack of effect cannot be attributed to the separation of onset and coda clusters into distinct sonority groups or to the separation of good vs. bad sonority profiles across syllable positions. For each of the children, the models that treat these four structures separately (i.e., the fully-crossed models) provided a better fit to the data than models that collapse sets of structures together.

**3.2 A closer look at Trevor’s onset clusters** The analyses in Figure 2 suggest that there is no systematic relationship between sequence and rate of acquisition when sets of clusters are considered. It is possible, however, that the expected relationship might emerge on a cluster-by-cluster basis. In other words, there might be effects associated with clusters’ specific segmental content that are obscured when multiple clusters are considered simultaneously.

To test this prediction Trevor’s onset target /stop + l/ and /stop + r/ clusters were isolated from the data above. These clusters are arguably ideal from the perspective of sonority sequencing (e.g., Clements 1990), and they are well represented among the set of rising-sonority onset clusters in Trevor’s data. Furthermore,

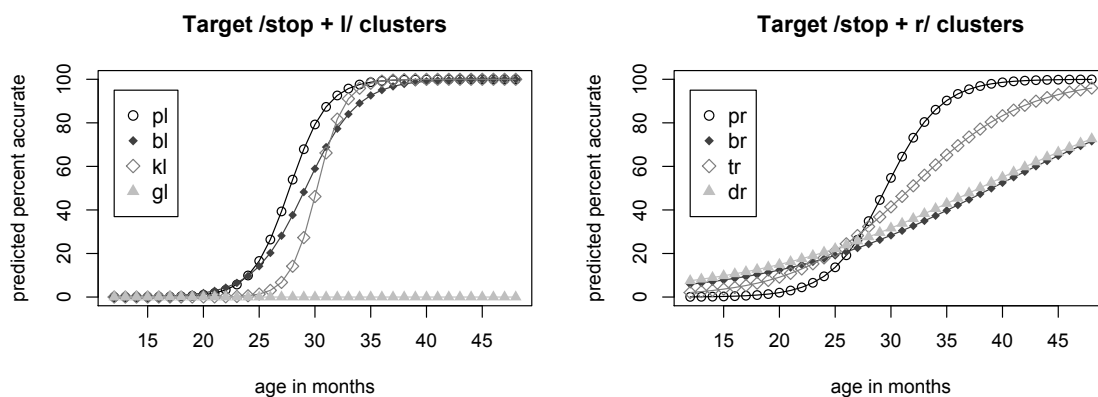
when these clusters are not realized accurately, they are all overwhelmingly repaired through deletion of the liquid consonant. Examples of these clusters and their realization by Trevor are given in (8).

|     |    |                            |              |        |                            |                   |         |
|-----|----|----------------------------|--------------|--------|----------------------------|-------------------|---------|
| (8) | a. | /stop + l/ target clusters |              | b.     | /stop + r/ target clusters |                   |         |
|     |    | 2;01.00                    | <i>play</i>  | [pe:]  | 1;11.09                    | <i>pretty</i>     | [pe:di] |
|     |    | 2;01.00                    | <i>blow</i>  | [bo:]  | 1;11.09                    | <i>brush</i>      | [bai:f] |
|     |    | 2;00.27                    | <i>clay</i>  | [ke:]  | 2;00.03                    | <i>Trevor</i>     | [te:wə] |
|     |    | 2;01.14                    | <i>glass</i> | [gæ:s] | 2;00.08                    | <i>dropped it</i> | [dæpɪt] |

Among Trevor's /stop + l/ target clusters, /pl/ was attempted 122 times, /bl/ was attempted 112 times, /kl/ was attempted 100 times, and /gl/ was attempted 9 times. Among his /stop + r/ target clusters, /pr/ was attempted 18 times, /br/ was attempted 82 times, /tr/ was attempted 177 times, and /dr/ was attempted 31 times.

A single logistic regression model was constructed for the set of eight target clusters. A model that included age (in months), target cluster type, and the interaction of age and cluster type as predictors provided a significantly better fit to the data than simpler models ( $p < 0.001$ ). The predictions of the best-fit model are plotted in Figure 3. Target /stop + l/ and target /stop + r/ clusters are plotted separately for readability.

Figure 3: Predicted accuracy rate by age in months for Trevor's /stop + l/ and /stop + r/ onset clusters



As these figures make clear, even at this most granular level of analysis, support for the expected relationship between sequence and rate of acquisition is not found. Among the target /stop + l/ clusters, for instance, accurate production of target /bl/ precedes accurate production of target /kl/, but the rate of acquisition for target /kl/ is faster than the rate of acquisition for target /bl/. This pattern is precisely the opposite of what the error-driven learning models predict. Similar patterns are evident among the target /stop + r/ clusters.

#### 4 Frequency and sequence of acquisition revisited

The absence of a clear relationship between sequence and rate of acquisition in these data raises important questions about the role of frequency within error-driven learning models. As sketched in §2, the predicted relationship between sequence and rate in these models is due to the key role played by frequency. The more frequently a target is attempted, the more frequently errors involving that target will be made, and the more rapidly the values of the constraints implicated in its realization will be adjusted. This means both that more frequent forms will begin to be realized accurately relatively earlier, and that they will proceed toward full accuracy at a relatively faster rate.

Most studies of learning paths (e.g., Boersma & Levelt 2000, 2003, Jarosz 2010) have taken the frequency of structures in child-directed speech as the basis for defining the input distribution to which the learning model is exposed. This approach is supported by a wide range of literature showing that children

show sensitivity to various aspect of the target language’s grammar even before first words emerge (e.g., de Boysson-Bardies & Vihman 1991).

Kirk & Demuth (2005) provide a baseline for the frequency of different cluster types in English child-directed speech based on the adult productions in the Brown (1973) and Bernstein-Ratner (1982) corpora. There, coda clusters comprised 67% of word-edge biconsonantal clusters and onset clusters comprised 33%. Kirk & Demuth’s (2005) counts included coda clusters formed through morphological concatenation, and so it is difficult to directly test the effect of coda vs. onset cluster frequencies against the children’s data in the current study. We can, however, look more closely at the onset domain, where morphological complexity is not a confound. There, Kirk & Demuth found that rising-sonority onset clusters were roughly 3.4 times more frequent than falling sonority onset clusters in child-directed speech. On this basis, we would expect all of the children to acquire rising-sonority onset clusters before, and at a faster rate than, falling-sonority onset clusters. The children’s attempts at the different cluster types provide a converging set of predictions. As reflected in (6), all of the children attempt rising-sonority onset clusters significantly more frequently than falling-sonority onset clusters (paired t-test  $t(5) = 5.183, p = 0.003$ ).

To test the role of frequency using the models from §3.1, it is necessary to operationalize the concepts “acquired first” and “acquired fastest.” For concreteness, we can say that the structure which is predicted to reach 5% accurate realization first is “acquired first.” “Acquired fastest” can then be operationalized as the structure which is predicted to move from 5% to 95% accuracy most quickly. As shown in (9), given these criteria, all six of the children show earlier acquisition of rising-sonority onset clusters than of falling-sonority onset clusters. There is no consistency, however, in whether rising- or falling-sonority onset clusters are acquired more quickly.

(9)

| Child  | Onset type acquired first | Onset type acquired fastest |
|--------|---------------------------|-----------------------------|
| Amahl  | rising sonority           | <i>falling sonority</i>     |
| Trevor | rising sonority           | rising sonority             |
| E.     | rising sonority           | rising sonority             |
| Alex   | rising sonority           | <i>falling sonority</i>     |
| Ethan  | rising sonority           | <i>falling sonority</i>     |
| Lily   | rising sonority           | <i>falling sonority</i>     |

From these data it appears that sequence of acquisition may be influenced by input frequency in the expected manner. It would be unwise however, to conclude that this provides robust support for the role played by frequency in the error-driven learning models identified above. First, these are data from only six children, and they involve only two types of structures, and so conclusions must necessarily be cautious. Second, rising-sonority onset clusters are also considerably less marked than falling-sonority onset clusters, and so it is not clear that frequency should be accorded a primary role in the explanation of this pattern (see Jarosz 2010 for related discussion). Finally, and most importantly, error-driven gradual learning models also predict that *rate* of acquisition should be influenced by frequency, which is clearly not the case here. It is not evident how the existing models could be revised to decouple the sequence and rate predictions without changing the basic workings of the system.

These results raise the question of how the current findings should be integrated with those of previous studies. There are at least two issues of relevance here. First, there is the question of determining what types of differences in frequency we can reasonably expect to find reflected in patterns of child language development. In the case of rising- vs. falling-sonority onset clusters, the frequencies in child-directed speech and in the children’s attempts are well aligned, and the difference in relative frequency between the two structure types is quite substantial. Similarly large differences in frequency feature in many of the studies cited in §1. Quantifying the differences in frequency necessary to see effects on sequence of acquisition is an important direction for future work.

The second issue to consider when comparing the current results with those of previous studies is the nature of the criteria used in determining when a given structure is “acquired.” Some studies consider a structure to be acquired once it has appeared a set number of times in the child’s speech (e.g., Levelt, Schiller & Levelt 1999/2000). This has the advantage that the identity of the target form need not be known in order for the criterion to be applied. On the other hand, this approach is confounded by the fact that if a

child does not attempt certain types of structure, they can never be deemed “acquired” – even if the child’s grammar would be able to reproduce the relevant forms accurately. This a particular concern to the extent that children’s attempt frequencies replicate the frequencies of forms in child-directed speech. The other possibility, then, is to adopt a definition of “acquired” that is based on a comparison of the target output and the child’s productions. This is the approach taken in (9) above.

Regardless of whether the input is taken into consideration, the threshold defined for “acquired” is crucial. Rate of learning interacts with sequence of acquisition in non-trivial ways. As we have seen, in some cases, structures that initially appear later may have a relatively fast learning rate, and so may reach high thresholds of accuracy earlier than structures with a relatively slower learning rate. This is evident even in the onset cluster data considered here. When the 5% accuracy threshold is applied, as in (9), all six of the children are deemed to have acquired rising-sonority clusters first. As (10) shows, however, the picture becomes less homogeneous when other accuracy thresholds are considered.

(10)

| Child  | First to reach 5% accuracy | First to reach 50% accuracy | First to reach 80% accuracy | First to reach 95% accuracy |
|--------|----------------------------|-----------------------------|-----------------------------|-----------------------------|
| Amahl  | rising sonority            | rising sonority             | rising sonority             | <i>falling sonority</i>     |
| Trevor | rising sonority            | rising sonority             | rising sonority             | rising sonority             |
| E.     | rising sonority            | rising sonority             | rising sonority             | rising sonority             |
| Alex   | rising sonority            | rising sonority             | <i>falling sonority</i>     | <i>falling sonority</i>     |
| Ethan  | rising sonority            | rising sonority             | <i>falling sonority</i>     | <i>falling sonority</i>     |
| Lily   | rising sonority            | rising sonority             | <i>falling sonority</i>     | <i>falling sonority</i>     |

Frequency may favour earlier acquisition of some structures, but it is clear that sequence of acquisition cannot be considered to be a single objective measure. Care is needed both in defining measures of frequency and in defining criteria for acquisition.

## 5 Conclusion

Many of the most commonly adopted error-driven gradual learning models predict a tight relationship between learning sequence and rate of acquisition. More frequent structures should be realized accurately earlier than less frequent structures. Then, once the initial accurate productions emerge, more frequent structures should progress toward full accuracy at a faster rate than less frequent structures. As a result, sequence of acquisition and rate of learning should be correlated, with structures that emerge earlier progressing toward full accuracy at a faster rate than structures that emerge later in the acquisition process. These predictions hold independent of the specific constraints involved.

Evidence from children’s acquisition of consonant clusters challenges the Rate-Sequence Prediction. Across the six children considered in this paper, learning sequence and rate of acquisition are largely disassociated. Some early-acquired structures show a rapid progression toward full accuracy, while others show a much slower progression. Furthermore, it is unclear that the expected role of frequency in favouring earlier and more rapid acquisition is manifested in these data. At best there is tentative support for the idea that earlier accurate realization may be possible for structures that are robustly favoured by both frequency and markedness considerations.

Overall, this paper points to the importance of carefully considering the predictions that learning models make for phonological acquisition, and of thoroughly assessing the extent to which these predictions are supported by child language data. More concretely, it suggests that many constraint-based models of error-driven gradual learning show a greater sensitivity to frequency than do children engaged in the actual task of phonological acquisition. Continuing research into formal models should aim to more accurately integrate frequency effects with the other factors relevant to the course of acquisition.

## Appendix

| <b>Amahl</b>                          | Estimate | S.E.    | z value | Pr(> z ) |
|---------------------------------------|----------|---------|---------|----------|
| (Intercept)                           | -10.6447 | 4.1888  | -2.541  | 0.011    |
| age in months                         | 0.3314   | 0.1270  | 2.610   | 0.009    |
| onset (vs. coda)                      | -43.9506 | 12.1541 | -3.616  | < 0.001  |
| good (vs. bad) sonority profile       | 4.3442   | 4.3224  | 1.005   | 0.315    |
| age in months × onset                 | 1.0637   | 0.3204  | 3.319   | < 0.001  |
| age in months × good sonority         | -0.1196  | 0.1313  | -0.911  | 0.362    |
| onset × good sonority                 | 42.7135  | 12.2226 | 3.495   | < 0.001  |
| age in months × onset × good sonority | -1.0474  | 0.3230  | -3.243  | 0.001    |

| <b>Trevor</b>                         | Estimate | S.E.   | z value | Pr(> z ) |
|---------------------------------------|----------|--------|---------|----------|
| (Intercept)                           | -19.7467 | 5.4986 | -3.591  | < 0.001  |
| age in months                         | 0.9946   | 0.2709 | 3.671   | < 0.001  |
| onset (vs. coda)                      | 13.2532  | 5.8535 | 2.264   | 0.024    |
| good (vs. bad) sonority profile       | 10.6612  | 5.5765 | 1.912   | 0.056    |
| age in months × onset                 | -0.8229  | 0.2826 | -2.912  | 0.004    |
| age in months × good sonority         | -0.5326  | 0.2747 | -1.939  | 0.053    |
| onset × good sonority                 | -12.9694 | 5.9682 | -2.173  | 0.030    |
| age in months × onset × good sonority | 0.6550   | 0.2875 | 2.278   | 0.023    |

| <b>E.</b>                             | Estimate | S.E.   | z value | Pr(> z ) |
|---------------------------------------|----------|--------|---------|----------|
| (Intercept)                           | -14.6261 | 7.2612 | -2.014  | 0.044    |
| age in months                         | 0.7521   | 0.3792 | 1.983   | 0.047    |
| onset (vs. coda)                      | 10.0622  | 7.5280 | 1.337   | 0.181    |
| good (vs. bad) sonority profile       | 10.5704  | 7.3848 | 1.431   | 0.152    |
| age in months × onset                 | -0.6114  | 0.3878 | -1.577  | 0.115    |
| age in months × good sonority         | -0.5823  | 0.3839 | -1.517  | 0.130    |
| onset × good sonority                 | -10.7105 | 7.7055 | -1.390  | 0.165    |
| age in months × onset × good sonority | 0.6171   | 0.3944 | 1.565   | 0.118    |

| <b>Alex</b>                           | Estimate | S.E.   | z value | Pr(> z ) |
|---------------------------------------|----------|--------|---------|----------|
| (Intercept)                           | -9.1697  | 1.3839 | -6.626  | < 0.001  |
| age in months                         | 0.3084   | 0.0441 | 6.987   | < 0.001  |
| onset (vs. coda)                      | 1.5099   | 2.3337 | 0.647   | 0.518    |
| good (vs. bad) sonority profile       | 1.8560   | 1.5669 | 1.185   | 0.236    |
| age in months × onset                 | -0.0502  | 0.0752 | -0.667  | 0.505    |
| age in months × good sonority         | -0.1023  | 0.0492 | -2.081  | 0.037    |
| onset × good sonority                 | 0.8862   | 2.4844 | 0.357   | 0.721    |
| age in months × onset × good sonority | 0.0188   | 0.0796 | 0.236   | 0.813    |

| <b>Ethan</b>                          | Estimate | S.E.   | z value | Pr(> z ) |
|---------------------------------------|----------|--------|---------|----------|
| (Intercept)                           | -0.5954  | 1.0690 | -0.557  | 0.578    |
| age in months                         | 0.0340   | 0.0475 | 0.715   | 0.475    |
| onset (vs. coda)                      | -7.1302  | 1.4863 | -4.797  | < 0.001  |
| good (vs. bad) sonority profile       | -2.1936  | 1.1476 | -1.911  | 0.056    |
| age in months × onset                 | 0.2412   | 0.0622 | 3.877   | < 0.001  |
| age in months × good sonority         | 0.0300   | 0.0500 | 0.601   | 0.548    |
| onset × good sonority                 | 6.3051   | 1.5742 | 4.005   | < 0.001  |
| age in months × onset × good sonority | -0.1546  | 0.0656 | -2.356  | 0.018    |

| Lily                                  | Estimate | S.E.   | z value | Pr(> z ) |
|---------------------------------------|----------|--------|---------|----------|
| (Intercept)                           | -0.6165  | 1.0867 | -0.567  | 0.571    |
| age in months                         | 0.0616   | 0.0351 | 1.755   | 0.079    |
| onset (vs. coda)                      | -1.4457  | 1.3348 | -1.083  | 0.279    |
| good (vs. bad) sonority profile       | -2.4076  | 1.1266 | -2.137  | 0.033    |
| age in months × onset                 | 0.0383   | 0.0439 | 0.874   | 0.382    |
| age in months × good sonority         | 0.0326   | 0.0363 | 0.899   | 0.368    |
| onset × good sonority                 | 4.2141   | 1.4018 | 3.006   | 0.003    |
| age in months × onset × good sonority | -0.0940  | 0.0460 | -2.042  | 0.041    |

## References

- Beckman, Mary E. & Jan Edwards. 2008. Generalizing over lexicons to predict consonant mastery. *Laboratory Phonology* 1: 319-343.
- Bernstein, Nan. 1982. *Acoustic Study of Mothers' Speech to Language-Learning Children: An Analysis of Vowel Articulatory Characteristics*. PhD dissertation, Boston University.
- Boersma, Paul. 1998. *Functional Phonology: Formalizing the Interaction Between Articulatory and Perceptual Drives*. The Hague: Holland Academic Graphics.
- Boersma, Paul & Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32: 45-86.
- Boersma, Paul & Clara Levelt. 2000. Gradual constraint-ranking learning algorithm predicts acquisition order. In E.V. Clark (ed.), *Proceedings of the 30th Child Language Research Forum*, 229-237. Stanford, CA: CSLI.
- Boersma, Paul & Claartje Levelt. 2003. Optimality Theory and phonological acquisition. *Annual Review of Language Acquisition* 3: 1-50.
- Boersma, Paul & Joe Pater. to appear. Convergence properties of a gradual learning algorithm for Harmonic Grammar. In John McCarthy & Joe Pater (eds.), *Harmonic Grammar and Harmonic Serialism*. London: Equinox Press.
- Boersma, Paul & David Weenink. 2012. *PRAAT: Doing Phonetics by Computer*. [www.praat.org].
- de Boysson-Bardies, Bénédicte & Marilyn May Vihman. 1991. Adaptation to language: evidence from babbling and first words in four languages. *Language* 67(2): 297-319.
- Brown, Roger. 1973. *A First Language: The Early Stages*. Cambridge, MA: Harvard University Press.
- Clements, G.N. 1990. The role of sonority in core syllabification. In John Kingston & Mary E. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, 283-333. Cambridge: Cambridge University Press.
- Compton, Arthur J. & Mary Streever. 1977. Child phonology: data collection and preliminary analyses. In *Papers and Reports in Child Language Development* 13. Stanford University.
- Curtin, Suzanne & Kie Zuraw. 2002. Explaining constraint demotion in a developing system. In Anna H.-J. Do, Laura Domínguez & Aimee Johansen (eds.), *BUCLD 26: Proceedings of the 26th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadia Press.
- Demuth, Katherine. 1995. Markedness and the development of prosodic structure. In Jill Beckman (ed.), *Proceedings of the 25th Meeting of the North East Linguistic Society*. Amherst, MA: GLSA. [ROA-50].
- Demuth, Katherine, Jennifer Culbertson & Jennifer Alter. 2006. Word-minimality, epenthesis and coda licensing in the early acquisition of English. *Language and Speech* 49(2): 137-174.
- Demuth, Katherine & Mark Johnson. 2003. Truncation to subminimal words in early French. *Canadian Journal of Linguistics* 48: 211 – 241.
- Edwards, Jan & Mary E. Beckman. 2008. Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in phonological development. *Language, Learning, and Development* 4: 122-156.
- Edwards, Jan, Mary E. Beckman & Benjamin Munson. 2015. Frequency effects in phonological acquisition. *Journal of Child Language* 42(2): 306-311.
- Gnanadesikan, Amalia. 2004. Markedness and faithfulness constraints in child phonology. In René Kager, Joe Pater & Wim Zonneveld (eds.), *Fixing Priorities: Constraints in Phonological Acquisition*, 73-108. Cambridge: Cambridge University Press.
- Goldwater, Sharon & Mark 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.
- Hayes, Bruce. 2004. Phonological acquisition in Optimality Theory: the early stages. In René Kager, Joe Pater & Wim Zonneveld (eds.), *Fixing Priorities: Constraints in Phonological Acquisition*, 158-203. Cambridge: Cambridge University Press.

- Ingram, David. 1989. *First Language Acquisition: Method, Description and Explanation*. Cambridge: Cambridge University Press.
- Inkelas, Sharon & Yvan Rose. 2003. Velar fronting revisited. In Barbara Beachley, Amanda Brown & Fran Conlin (eds.), *Proceedings of the 27th Annual Boston University Conference on Language Development*, 334-345. Somerville, MA: Cascadilla Press.
- Inkelas, Sharon & Yvan Rose. 2008. Positional neutralization: a case study from child language. *Language* 83: 707-736.
- Jarosz, Gaja. 2010. Implicational markedness and frequency in constraint-based computational models of phonological learning. *Journal of Child Language* 37: 565-606.
- Jesney, Karen & Anne-Michelle Tessier. 2011. Biases in Harmonic Grammar: the road to restrictive learning. *Natural Language and Linguistic Theory* 29(1): 251-290.
- Kirk, Cecilia & Katherine Demuth. 2005. Asymmetries in the acquisition of word-initial and word-final consonant clusters. *Journal of Child Language* 32: 709-734.
- Levelt, Clara C., Niels O. Schiller & Willem J. Levelt. 1999/2000. The acquisition of syllable types. *Language Acquisition* 8(3): 237-264.
- MacWhinney, Brian. 2000. *The CHILDES project: Tools for analyzing talk. Third Edition*. Mahwah, NJ: Lawrence Erlbaum.
- Menn, Lise. 2004. Saving the baby: making sure that old data survive new theories. In René Kager, Joe Pater & Wim Zonneveld (eds.), *Fixing Priorities: Constraints in Phonological Acquisition*, 54-72. Cambridge: Cambridge University Press.
- Pater, Joe. 1997. Minimal violation and phonological development. *Language Acquisition* 6(3): 201-253.
- R Core Team. 2014. *R: A Language and Environment for Statistical Computing*. [www.R-project.org].
- Roark, Brian & Katherine Demuth. 2000. Prosodic constraints and the learner's environment: a corpus study. In S.C. Howell, S.A. Fisch & T. Keith-Lucas (eds.), *Proceedings of the 24th Annual Boston University Conference on Language Development*, 597-608. Somerville, MA: Cascadilla Press.
- Smith, Neilson V. 1973. *The Acquisition of Phonology: A Case Study*. Cambridge: Cambridge University Press.
- Smolensky, Paul. 1996. On the comprehension / production dilemma in child language. *Linguistic Inquiry* 21: 720-731.
- Tessier, Anne-Michelle. 2007. *Biases and Stages in Phonological Acquisition*. PhD dissertation. University of Massachusetts Amherst.