Michael Becker (becker@fas.harvard.edu)
Anne-Michelle Tessier (amtessier@ualberta.ca) May 2, 2011

Trajectories of faithfulness in child-specific phonology*

1 Introduction

This paper describes, contrasts and analyzes two kinds of developmental paths in the phonological acquisition of Trevor, a child learning North American English (Compton & Streeter 1977; Pater 1997b). Using longitudinal data (ages 0;10;11 to 3;1.8), we demonstrate that Trevor acquired onset consonant clusters and words with multiple consonantal places of articulation along two different trajectories: faithfulness to onset clusters developed along an S-shape path, increasing steadily over time, while faithfulness to some combinations of multiple places of articulation showed a U-shaped pattern. We start by establishing the statistic significance of the U-shape pattern, and then offer a computationally-implemented analysis of both developmental paths that expands on the Optimality-Theoretic learning models in Prince & Tesar (2004); Hayes (2004); Tessier (2007, 2009). Both S-shaped and U-shaped learning will be shown to emerge from the interaction of an OT grammar, a constraint demotion algorithm that favors ranking Markedness constraints above Faithfulness, a stored Cache of previously-produced forms, and a "dual route" mechanism for producing variable outputs over the course of acquisition.

One broad issue that this paper addresses is the sources of gradual, incremental change and associated variation in child phonological acquisition. It is an uncontested fact that children's successive stages of temporarily-stable phonological production are often overlapping, moving from one to the other with intermediate

*For their thoughtful comments and questions, we thank Adam Abright, Karen Jesney, Joe Pater, John McCarthy, Yvan Rose, three anonymous reviewers, an associate editor, and audiences at the University of Alberta, the University of Massachusetts Amherst, and the 84th LSA meeting in Baltimore. We are indebted to Peter Graff and Corey Telfer for their advice on statistical methods.

periods of variability. Particularly clear examples come from longitudinal studies such as Smith (1973); Menn (1976); Fikkert (1994).

The literature on phonological acquisition, however, is in much less agreement as to the import of these intermediate periods. Many recent constraint-based learning procedures propose that variability be captured by grammatical mechanisms. In this case, a grammar may produce multiple outputs for a single input at any one stage, and the frequency with which each output occurs at a particular stage varies as a function of grammatical properties, such as the relative proximity of constraints along a number line or a more direct, grammatically-assigned probability. Such works include Boersma (1997, 1998); Boersma & Levelt (2000); Boersma & Hayes (2001); Curtin & Zuraw (2001); Goldwater & Johnson (2003); Jarosz (2006); Hayes & Wilson (2008); Jesney (2007).

On the other hand, some earlier views of child phonology understand children's variability, and equally their development through successive stages of non-target phonologies, as the result of multiple representations for a given word — or perhaps entire multiple lexicons, which store both the observed adult forms and the child's current productions (see especially Menn 1983; Menn & Matthei 1992; for related proposals see also Ingram 1974, Dinnsen & Elbert 1984, also Ferguson & Farwell 1975; Macken & Ferguson 1981, and Bernhardt & Stemberger 1998:43–55.)[1]

In a way, this paper's proposals form a hybrid of these two types of accounts.[2] At its heart is a "classic" OT grammar (Prince & Smolensky 1993/2004; Tesar & Smolensky 2000; Moreton 2004), meaning that at every point in time the learning procedure entertains a single constraint ranking, which in the simplest case will always choose a single optimal output candidate. Nevertheless, our learning procedure (henceforth, the *learner*)[3] produces gradual change, relying both on stored child-specific forms as well as online grammar use, and its incremental changes come from its multiple methods of production rather than a variable grammar. This paper therefore offers a way to learn using re-ranking algorithms

[1] A further view is that such variation is created solely by performance factors, or taken as evidence that child production has nothing to do with grammar (Hale & Reiss 1998).

[2] We are grateful to an anonymous reviewer in particular for illuminating the connections between our proposal's use of stored errors and previous accounts that used multiple child lexicons, particularly Menn (1983).

[3] See §3 for a discussion of the multiple related meanings of the term "learner".

that operate on classic OT grammars (Prince & Tesar 2004; Hayes 2004), and thus retain their benefits, while also capturing some dimensions of realism in natural language learning.

Before continuing, we note that there is obviously variation in target grammars too, in which multiple optimal outputs are necessary. Given the known empirical differences between the developmental variability that we study here and the variability observed in adult grammars, we take this to be a strength of our approach. We decouple the variation associated with the development of a learner's grammar from variation created by a mature adult grammar — a point we revisit in §5.

This paper also deals with larger questions about the nature of OT's constraint set, CON, because it crucially assumes that some markedness constraints appear in the learners' grammar *during the course of learning*, while other constraints are present at the earliest production stages. While the standard hypothesis in the OT literature is that all of CON is pre-made — and that languages differ solely on the ranking of constraints, not their number or content — we note that several studies on child phonology have argued that children induce some of their own markedness constraints in the learning process: see for example Pater (1997a); Goad (2001); Levelt & van Oostendorp (2007); Fikkert & Levelt (2008); Inkelas & Rose (2008). We do not propose an explicit mechanism for how learners induce markedness constraints, for which the reader is referred to Hayes (1999); Flack (2007); Hayes & Wilson (2008); Boersma & Pater (2007); Moreton (2010) — what we propose here is a way to diagnose the effect of induction, if induction occurs late enough in the learning process. At Trevor's earliest stages, onset clusters are consistently simplified because *ComplexOnset is present in the top stratum at the initial stage, $\mathcal{H}_0$; at the same time, the earliest words remain faithful to consonantal place (i.e. resisting consonant harmony) because Agree(KVT) is not yet present at $\mathcal{H}_0$. In this system, U-shaped development can only arise when a new markedness constraint is added to CON.

The rest of this paper is organized as follows. Section 2 presents the data of Trevor's complex onset simplification and consonant harmony, and then the statistical models which support two different paths towards their respective acquisition.

Section 3 presents our learner, which builds on Tessier's (2009) Error Selective Learning model, and demonstrates how this learner's dual-route component continually puts the current grammar in competition with stored productions and so helps create variability of both S-shaped and U-shaped varieties. Section 4 shows how this learner produces both of Trevor's patterns of faithfulness, using a computational implementation. Section 5 discusses our results, and returns to the broad issues raised above.

2 The data

The first two subsections below present the longitudinal facts of Trevor's onset cluster simplification (§2.1) and consonant harmony (§2.2), drawn from the corpus of Compton & Streeter (1977); Pater (1997b). We then highlight the novelty of these results (§2.3) and present our statistical support for treating some consonant harmony patterns as significantly U-shaped (§2.4).

The Trevor corpus includes productions from ages 0;10;11 to 3;1.8 recorded on 187 different days. Thus, the corpus offers regular and representative samples of Trevor's development, but does not include every single word he said.

2.1 Faithfulness to complex onsets

English has a fairly large range of onset clusters, mostly obstruent-liquid and sC clusters. Like most children learning English, Trevor's early stages of production were consistently unfaithful to such clusters, most often by deleting all but one of the segments. This pattern is exemplifed in (1) below. For similar patterns in other L1 child phonologies, see e.g. Smith (1973); Gnanadesikan (2004); Pater & Barlow (2003); Rose (2000); Fikkert (1994).

In its intolerance to complex onsets, Trevor's early phonology resembles those of many adult target languages. For example, Egyptian Arabic repairs complex onsets by epenthesis (/t-rama/ → [ʔɪt.ra.ma], *[tra.ma] 'PASSIVE-to.throw', Kiparsky 2002 and references within), while many varieties of Finnish repair complex onsets by

deletion (e.g. *stressi* → *ressi* 'stress', Suomi et al. 2008:56).

(1) Trevor simplifies complex onsets[4]

type	target	productions	
PR	fɹɑg	hɑg (1;8), hʌwɑg (1;10)	'frog'
	pleɪ	peɪ (2;1)	'play'
TR	tɹi	ti (1;9), ʧi (1;4), tɛɹi (1;10)	'tree'
	dɹaɪv	daɪv (1;11)	'drive'
KR	klɑk	kæk, kɑ (1;1)	'clock'
	kwɪk	kɪk (2;0)	'quick'
sC	sniz	niz, əniz (1;7)	'sneeze'
	stoʊv	doʊv (1;11)	'stove'
	splæʃ	bæʃ (2;1)	'splash'

Trevor's repair of choice is deletion, which occurs in 52% of all tokens and 92% of unfaithful tokens (2). Other repairs include epenthesis (e.g. *fɹag*→ *hʌwag* 'frog'), vocalization (e.g. *sniz* → *əniz 'sneeze'*), and metathesis (e.g. *tɹɛvəɹ* → *tɛɹvɛɹ* 'Trevor').

(2) Trevor's repairs for complex onsets

	PR	TR	KR	sC	total	
deletion	530	232	266	260	1288	52%
faithful	427	347	108	197	1079	44%
epenthesis	15	17	11	12	55	2%
vocalization	7	1	0	34	42	2%
metathesis	0	5	2	0	7	<1%

The four panels in Figure 1 show the development of Trevor's onsets, divided by cluster type. We see essentially the same picture in all four: an initial period of uniformly simplified clusters, i.e. 0% faithfulness, slowly gives way to a period of variation, followed by an adult-like period of close to 100% faithfulness.[5]

[4]Here and throughout, P, T and K stand for all labial, coronal and dorsal consonants, R stands

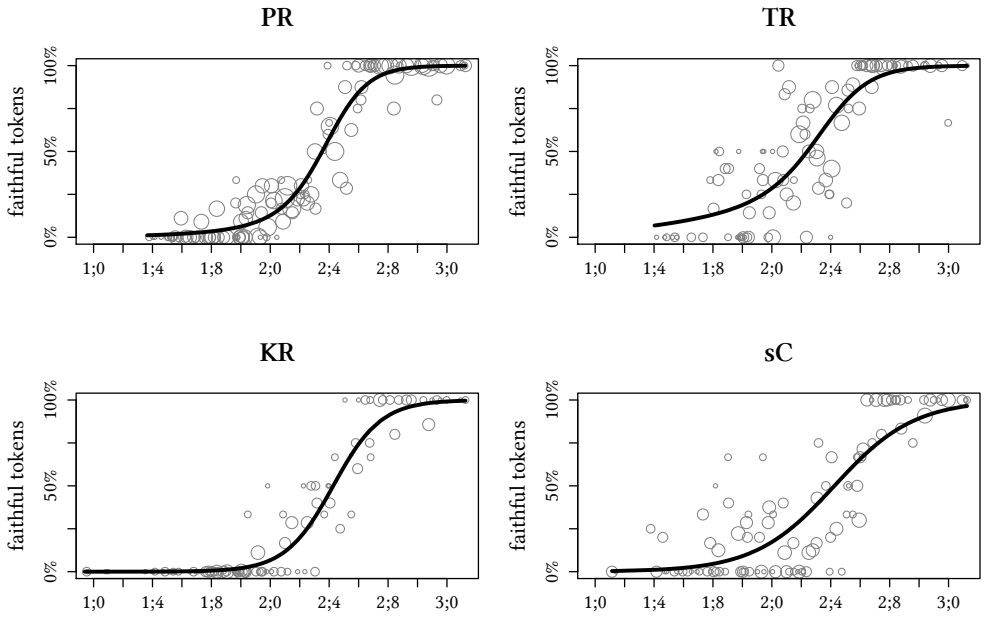


Figure 1: Monotonically increasing faithfulness to targets with complex onsets. In these panels, each bubble represents the proportion of faithful tokens on a given day of recording, with the size of the bubble proportional to the number of tokens (ranging from 1–30); the solid lines represent the predictions of a regression model, as explained in §2.4 below.

These panels all show a classic developmental path, going from an initial period of unmarked productions to a period of adult-like productions, with an intermediate period of variation. We stress that the variation is observed at every level of analysis, down to single types. For example, Trevor simplified the word *play* 19 times when he was 1;9.19–2;1.5. Then he pronounced 5 faithful tokens and 36 simplified tokens when he was 2;1.14–2;4.24. From 2;5.4 and onwards, he is recorded saying *play* faithfully 54 times. The point is that a gradual and variable transition from simplified complex onsets to faithful complex onsets is a necessary part of the analysis, and not an artifact of lumping different clusters together as we have done in Figure 1.

for all sonorants, C stands for any consonant, and V stands for any vowel.

[5]In the Trevor corpus, only 11% of records from 2;6.7 onwards contain a phonetic transcription, compared to 99.7% of the records before 2;6.7. While this transition surely indicates the transcribers' fatigue, it also indicates Trevor's approximation of adult phonology. Non-adult-like productions were still transcribed, as evidenced by the 11% of the records that were transcribed. We have simplified somewhat by counting untranscribed records as uniformly faithful; we suppose that in reality most, but not all, untranscribed tokens were produced faithfully.

2.2 Faithfulness to consonantal disharmony

Trevor's early phonology is again similar to many other L1 children's in that it often imposes harmony among consonantal place of articulation – for example, replacing a coronal consonant with a velar one when another velar follows, producing *duck* as [gʌk]. Unlike the complex onset simplification of the previous section, it has been frequently noted that long-distance consonant place harmony of this sort is never attested in adult languages (Vihman 1978; Pater 1997b; Pater & Werle 2001; Hansson 2001); this will be crucial to our analysis in section 3.

Representative examples of Trevor's consonant harmony are below in (3), including various combinations of place features with varying dominant consonants. For some samples of consonant harmony in other child phonologies, see e.g. Vihman (1978); Smith (1973); Menn (1983); Goad (1997); Rose (2000); Levelt (1994); Bat-El (2010); for discussion of Trevor's consonant harmony in particular, see Pater (1997b); Pater & Werle (2001, 2003).

For reasons that will become clear below, we refer to harmonic productions by the place of articulation that surfaces: thus a production of *duck* as [gʌk] is K-dominant, and a production of *duck* as [dʌtə] is T-dominant.

(3) Examples of Trevor's consonant harmony

places	adult target		Trevor's production	
KVT	goʊt	'goat'	goʊk (1;5)	K-dominant
	kæt	'cat'	tæt (2;0)	T-dominant
TVK	dɔg	'dog'	gɔg (1;7)	K-dominant
	dʌk	'duck'	dʌtə (1;3)	T-dominant
TVP	tɑp	'top'	pʌp (1;5)	P-dominant
	tɑm	'Tom'	tɑn (1;5)	T-dominant

While our main focus in this section is on consonant harmony, we must also address other relevant patterns of unfaithfulness in Trevor's speech. In table (4) and thereafter, we use the term *unfaithful* to only describe changes in major place of articulation; other changes are labeled *faithful*, including devoicing (e.g. [bɛːt]

for ‘bed’, 1;10), stopping of fricatives (e.g. [gɪːb] for ‘give’, 2;1), spirantization of affricates (e.g. [kɪːʃɪn] for ‘kitchen’ 2;7) and minor place changes (e.g. [fɪs] for ‘fish’, 2;3). Trevor does use coda deletion, which mostly affects [p] and [t]. We know that this deletion is not *driven* by consonant harmony, because (a) we get deletion in words with only one stop, such as *up*, and (b) deletion interacts opaquely with consonant harmony, giving rise to productions such as [gʊ] ‘book’ (1;5) and [dæːɪtə] ‘get it’ (2;0). Deletion mostly effects Trevor’s earliest words, such as *cup* and *dog*, which raises the possibility that his input didn’t contain these final stops. Thus the unfaithful forms reported in table (4) and thereafter are limited to those tokens where Trevor gives evidence that his input contained two consonants, either because he pronounced both, or because the surfacing single consonant harmonized with the other.

We note that Trevor’s unfaithfulness to velar place of articulation as in [dæːɪtə] for ‘get it’ (2;0), driven by harmony, contrasts with that of some children who replace velar consonants with coronal ones regardless of featural context (velar fronting, e.g. Chiat 1983; Inkelas & Rose 2008). Trevor is generally faithful to initial velars, as seen in words like *cow, cows, go, goes*. Even in the one word that seems to display fronting ([ki] → [ti] ‘key’), we see T-dominant productions significantly more often in the presence of the coronal plural suffix: only 2 out of 34 tokens of [ki] front to [ti] vs. 6 out of 18 tokens of [kiz] that front to [tiz] (Fisher exact test, $p < .05$).

Table (4) reports Trevor’s faithful and unfaithful tokens of disharmonic inputs, by place of articulation. The category “other” includes those rare examples in which the surface place appears to be unpredictable from the input (e.g. four tokens of [kʌtə] for ‘cup’, 1;2–1;3).

(4) Trevor’s treatment of disharmonic words

	KVP	KVT	PVK	PVT	TVK	TVP	total
faithful	510	906	493	1349	381	429	4068
K-dominant	9	38	194	0	186	0	427
P-dominant	0	0	0	191	0	19	210
T-dominant	0	90	0	4	30	5	129
other	5	0	0	1	0	8	14

Our central claim in this section is that Trevor's treament of some potentially-harmonic words – in particular, KVT inputs – does not follow an S-shape path towards faithfulness. Instead, KVT words show a U-shaped trajectory: first highly faithful, then *decreasingly so*, and then eventually increasingly faithful until target-like. This U-shape in accuracy has been noted in numerous areas of linguistic development, perhaps most often in morpho-phonological development, but in many other places as well; see e.g. Leopold (1939-1947); Bowerman (1982); MacWhinney (1978); Clahsen (1999); Bernhardt & Stemberger (1998) and other references below. We start here by characterizing the trajectory of consonant harmony in Trevor's lexicon as a whole, and then hone in on KVT words in particular.

The claim that Trevor applies different harmonies rather differently across time is illustrated by Figure 2 below. The three panels on the left show a familiar S-shape for faithfulness to TVK, PVK, and PVT words, beginning with a period of uniformly unfaithful productions (∼1;0–1;4 for TVK, ∼1;0–1;7 for PVK, ∼1;1–1;4 for PVT) and then gradually moving towards largely faithful productions.

In contrast, the three panels on the right for KVT, KVP, and TVP words show no sign of an S-shape; instead, these words appear relatively faithful throughout. In fact, the KVP and TVP panels hardly show any unfaithful productions at all (9 and 24 unfaithful tokens, which are 2% and 5% of the tokens on those panels, respectively). We therefore set them aside.

Unlike the three S-shape panels, and the two consistently-faithful panels, the KVT panel is of the most interest here. These targets follow two different trajectories: after an initial period of largely *faithful* productions from ∼1;0–1;5 (27 faithful tokens vs. 5 unfaithful ones), there is a long period of decline in faithfulness, and a rise into adult-like speech. Following the practice in the literature, we use the term U-shaped development for this down-and-up change.

We take a closer look at the KVT panel in Figure 3, where it is shown with its mirror image, the TVK panel. The U-shape is now seen to have further internal structure: KVT's initial faithful period gives way to a period of K-dominant harmony (e.g. *kɪs* → *kɪk* 'kiss', *kɪɾi* → *kiki* 'kitty'), which gradually transitions into a period of T-dominant harmony (e.g. *kɪs* → *tɪs, kɪɾi* → *tidi*). The faithful productions return

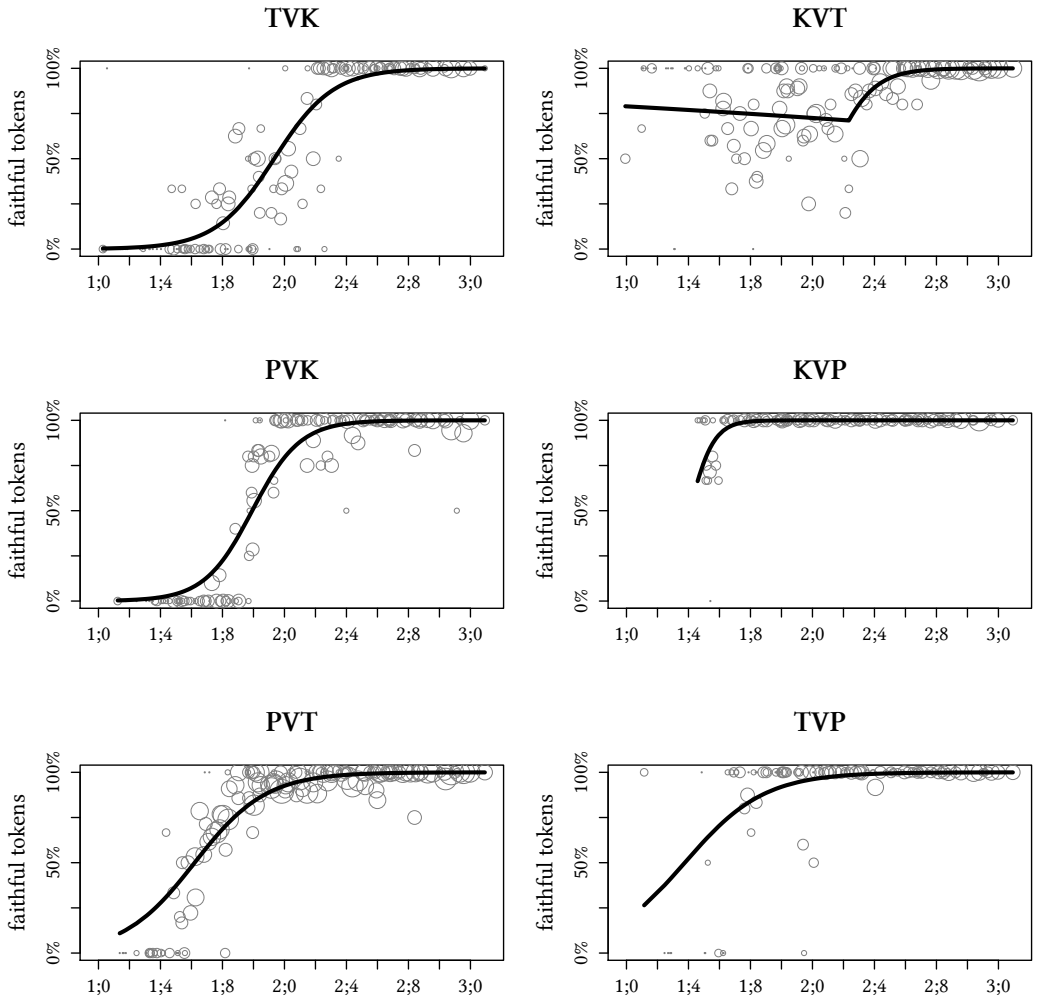


Figure 2: Development of resistance to consonant harmony by place of articulation

at ∼2;7.

Interestingly, the TVK panel shows the same transition from K-dominance to T-dominance, and at around the same time — circa 1;8 in both panels. The central difference between the two panels is that K-dominant harmonic outputs are strongly attested in the TVK panel from the very beginning, whereas there are very few harmonic tokens in the beginning of the KVT panel. We also note that the K-dominant period persists into the T-dominant period in the TVK panel, whereas the two periods show more separation in the KVT panel.

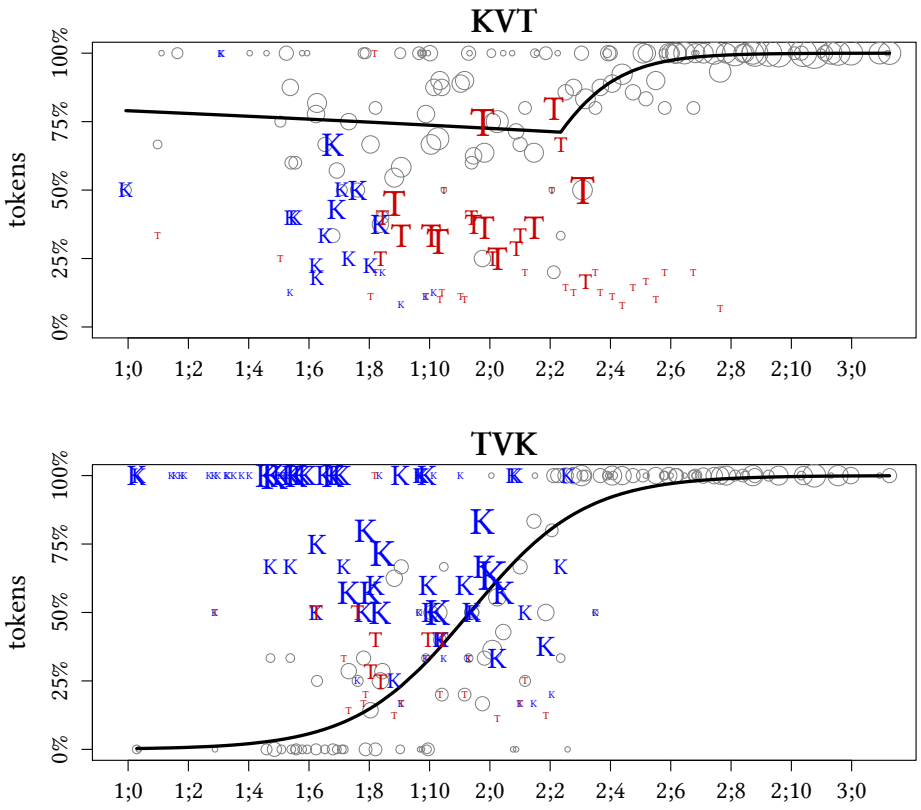


Figure 3: Direction of consonant harmony in KVT and TVK words. In these panels, each K or T represents harmonic tokens in a day of Trevor's productions, with the size of the letter proportional to the number of harmonic tokens recorded.

2.3 The U-shape in phonological development

This paper is the first, to our knowledge, to substantiate claims about phonological U-shaped development quantitatively. While there are numerous reports about U-shaped development (see e.g. Bernhardt & Stemberger 1998: 263–268 and references therein), most have not been able to draw on sufficiently rich longitudinal records or a sufficiently wide lexicon to allow a statistical analysis. In fact, most phonological U-shapes discussed in the literature document the development of an exceptional set of words, rather than an entire phonological process itself. For example, Leopold (1939-1947) reports the famous trajectory of the word 'pretty' in the speech of H: produced approximately correct at roughly 0;10–1;9, and then decreased in faithfulness to [pɪti] and later [bɪɾi]. Leopold noted that 9 other words

followed a similar trajectory during the first two years of H's speech, but this study was not designed to determine what proportion of H's outputs these 9 or 10 words represented, compared either to the entire lexicon or the relevant subsets. Similar examples of lexically-restricted U-shapes[6] are found in Moskovitz (1973); Bleile & Tomblim (1991); Menn (1983); Macken & Ferguson (1981). Cases of phonological U-shapes that are not necessarily lexically-restricted, and which show considerable variation and multiple regressions, are also reported in Stemberger & Bernhardt (2001); Stemberger et al. (2001); Ullrich et al. (2008). We also note that in the context of morphophonological development, much work has reported U-shaped patterns in the application of regular vs. irregular morphology (e.g. 'feet' → 'foots' → 'feet'), including longitudinal corpus studies (see for example Marcus et al. 1992; Maratsos 2000); see also the interesting U-shaped morpho-phonology reported in Kazazis (1969).

2.4 Statistical analysis

Figures 1, 2, and 3 above showed developmental trajectories using smooth continuous lines that were drawn through rather noisy data. In this section, we explain and justify the statistical analysis that generated these smooth lines. Specifically, we set out to demonstrate two main technical points:

(5) The U-shaped development seen in KVT words is significantly different from S-shaped development, even with the rather small number of tokens in the data.

(6) The U-shaped development is not attributable to a small number of atypical types; it is characteristic of Trevor's development in general.

We start with a logistic regression analysis, using the *glm()* function in R (R Development Core Team 2007).[7] In all of our models, the dependent variable is a binary factor: faithful or unfaithful production. The predictor is Trevor's age in days. A base model that only contains this predictor is limited to describing a

[6]Or at least, regressions from earlier faithful productions (called 'progressive idioms' by Moskovitz 1973) to later less accurate forms, which must have returned to faithfulness eventually.

[7]An R script for the analysis can be found at `http://becker.phonologist.org/trevor/`

linear correlation between age and faithfulness: Faithfulness can go up with age (as is observed e.g. in the TVK panel), or, in principle, faithfulness could go down with age, describing a child who starts well but then becomes less adult-like in their speech.[8]

Crucially, the base linear model is unable to describe faithfulness that goes down with age and then up again, or vice versa. To allow down-and-up development, we add a change point (or break point), which allows faithfulness to change its trajectory at one point in time (see e.g. Baayen 2008 §6.4, Jaynes 2003 §15.8.1, Mudelsee 2010 §4.2.2). The location of this change point is determined by considering each value on the x-axis (representing a day in Trevor's corpus) as a potential change point, and choosing that point which most closely fits the data (as measured by ANOVA model comparison). Adding a change point is one of several different approaches to non-linearity in statistics, and has the advantage of being a simple one. Other approaches include the use of restrictive cubic splines, and polynomials; for a Bayesian approach to a similar problem, see Steyvers & Brown (2006). The effect of the added change point is shown in Figure 4, where the TVK and KVT panels are repeated from Figure 2. The dashed lines show the predictions of the base linear models, while the solid lines show the predictions of the models that are augmented with change points. In the TVK panel, both models follow a trajectory of increasing faithfulness, and offer qualitatively similar descriptions. In the KVT panel, the base model is limited to showing gradual increase in faithfulness, and thus offers a poor fit to the data; the change point model describes the observed trend more closely.[9]

The same pattern holds when the KVT productions are separated according to the place of articulation of the output. Figure 5 shows only tokens of KVT words that were produced either faithfully or with K-dominant harmony. Again, the change

[8]The correlation is strictly linear when we calculate the logarithm of the odds of a faithful production. The straight line turns into an S-shape when we turn the logarithm of the odds into probabilities, which are limited to the interval between 0 and 1.

[9]The change point models make a significant improvement over the respective base linear models in both panels, as determined by an ANOVA model comparison, though the improvement is much bigger in the KVT panel (TVK: $\chi^2(1) = 25.34$, $p < .0001$, KVT: $\chi^2(1) = 61.61$, $p < .0001$). The qualitative difference between the two panels is that in the TVK panel, the trajectory of faithfulness is positive both before and after the change point, making only a change in the rate of increase in faithfulness; in the KVT model, the trajectory of faithfulness is negative before the change point and positive after it, effectively describing the down-and-up U-shape that is of interest.

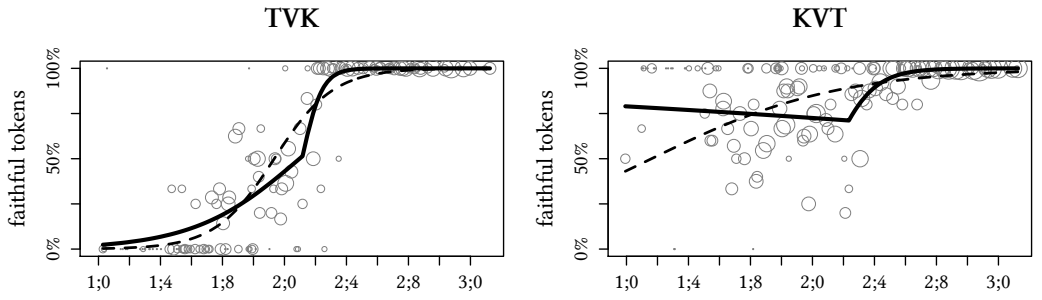


Figure 4: Base linear model (dashed) vs. lineal model with change point (solid)

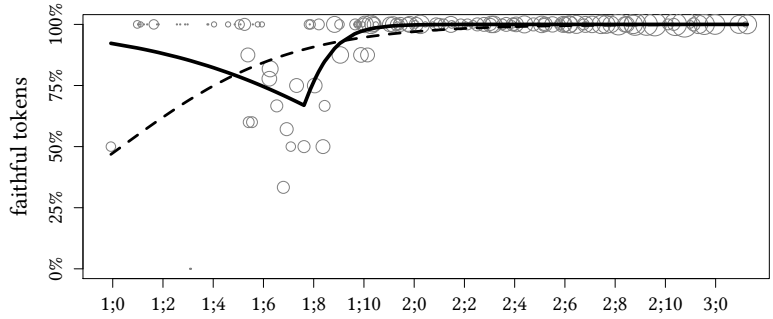


Figure 5: K-dominant harmony for KVT targets: base linear model (dashed) vs. change point model (solid)

point model (solid line) fits the data much better than the strictly linear model (dashed line).

The change point models provide a significantly better fit to the data than the strictly linear ones, as shown in an ANOVA model comparison.[10] In the two models shown in Figure 5, the improvement is highly significant ($\chi^2(1) = 38.54$, $p < .0001$). AIC (Akaike's Information Criterion, another method of model comparison) is similarly improved from 137.26 to 100.72.

The presentation so far relies solely on tokens, yet there is every reason to believe that types too should be considered by the model. Trevor's words differ widely in their token frequency and their faithfulness, and the overall pattern that we

[10]Following common practices in statistics (Crawley 2007; Baayen 2008) we use ANOVA as a general measure of data likelihood, here in conjunction with the χ^2 test, which is suited to the count data we have. A more specific use of ANOVA in conjunction with the F test is standard in the psycholinguistic literature.

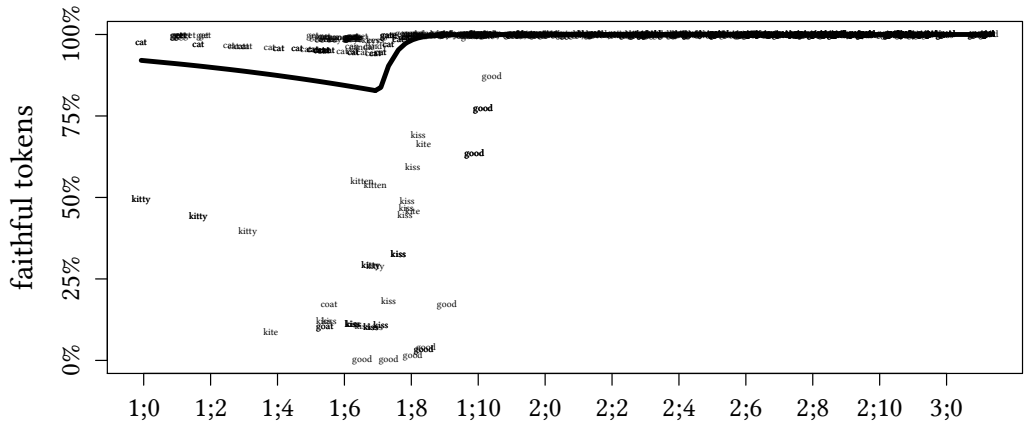


Figure 6: Hierarchical model predictions for K-dominant harmony of KVT targets

attribute to the grammar is generalized from the behavior of these individual words. To inform the statistical model about the classification of tokens into types, we switch from classical regressions to hierarchical regressions (also known as mixed-effects regressions), which we model using the *lmer()* function of the *lme4* package (Bates & Maechler 2009) in R. By using hierarchical regressions, we can group together all tokens according to their types, and the model can estimate levels of faithfulness taking types into account.

Figure 6 shows the same progression of K-dominant harmony as in Figure 5, but this time showing the separation of types in its predictions: some types are predicted to have a very low dip in faithfulnesss, and others much less so. The average prediction, i.e. a prediction for a new word if Trevor had an additional one, is shown with the solid line. While the predicted dip in faithfulness is not dramatic (83% faithfulness at 1;6.29), the improvement of the change point model over the base linear model is nevertheless highly significant (ANOVA model comparison, $\chi^2(1) = 25.26,\ p < .0001$). AIC is improved from 131.79 to 108.54, BIC (Bayesian Information Criterion) is improved from 146.34 to 127.94.

Similar results obtain when we examine the trajectory of T-dominant harmony, as shown in Figure 7. The solid line again shows the prediction for an average type, or a new word, if Trevor had one. The dip in faithfulness is not very pronounced (90% faithfulness at 2;2.7), but again the improvement of the change point model over

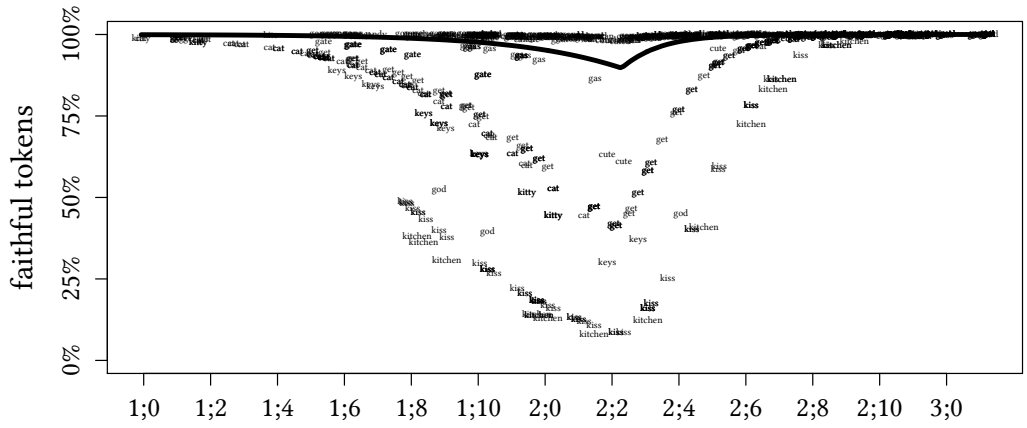


Figure 7: Hierarchical model predictions for T-dominant harmony of KVT targets

the base linear model is highly significant (ANOVA model comparison, $\chi^2(1) = 91.45$, $p < .0001$). AIC is improved from 406.95 to 317.49, and BIC from 421.66 to 337.11.

Pater & Werle (2001) report that Trevor is more likely to harmonize dorsal and coronal consonants across a back vowel than across a front vowel, and that in general harmony is as or more likely when crossing a homorganic vowel. (For similar patterns see also Levelt 1994 on Dutch and Stoel-Gammon 1996 on a different English example.) Our statistical evaluation of Trevor's corpus indeed reveals a front/back connection between harmonizing consonants and vowels: T-dominant harmony is significantly more likely with an intervening front vowel, and K-dominant harmony is significantly more likely with an intervening back vowel. Since these vowel effects offer only modest improvements to the models (T-dominant: $\chi^2(1) = 6.3$, $p < .05$; K-dominant: $\chi^2(1) = 4.51$, $p < .05$) we leave aside vowel quality for the remainder of this study, but see Pater & Werle's (2001) treatment of vowel influences in Trevor's harmony and cf. Fikkert & Levelt (2008).

3 An analysis of gradual, variable learning

The goal of our analysis is to equip a learner whose constraint ranking is inherently categorical with the mechanisms to learn gradually and variably over time. The class of grammars being learned here are ordinal ("classic") OT rankings, i.e. where every constraint is ranked above or below every other, rather than stochastically (cf. Boersma 1997; Boersma & Hayes 2001) and where winners are determined via strict domination (Prince & Smolensky 1993/2004) rather than a harmony score (cf. Smolensky & Legendre 2006). We start with a very brief introduction to constraint demotion algorithms, and present our basic view of how the learner uses them in §3.1. We apply our learner to a simple case of S-shaped development using Trevor's complex onsets in §3.2. Then, we show how the introduction of a new markedness constraint causes U-shaped development by analysing Trevor's KVT words in §3.3. Finally, we integrate the S-shaped development of TVK words in §3.4, for a fuller analysis of Trevor's consonant harmony.

Our approach begins with the algorithms for ordinal OT learning proposed by Prince & Tesar (2004) and Hayes (2004), which both aim to ensure that learners map from stored errors to new grammars as restrictively as possible — that is, by building rankings that can faithfully reproduce the surface forms they observe, but otherwise produce as few unheard surface forms as possible. When building new grammars from sets of errors, the learner proceeds by "resolving" errors – that is, by installing constraints in an order that ensures the next grammar will prefer the target observed form (the "winner") to the error created by the last grammar (the "loser").

At the earliest stage, the ranking algorithm places all Markedness constraints above all Faithfulness constraints, as in the grammar fragment in (7) below, and this initial ranking can make an error like (8), where /fɪag/ comes out as [fag].

The error format in (8) compares the winner and loser forms on the merits of each constraint (Prince 2002): assigning an L to *ComplexOnset which prefers the Loser's singleton onset, while assigning a W to Max which prefers the Winner's preservation of the cluster. From this error, the learner can build the improved ranking in (9), which no longer simplifies complex onsets (because Max ranks

above *ComplexOnset) but otherwise retains its ban on other marked structure (because in the absence of positive evidence, other markedness constraints like *ComplexCoda remain above Max).

(7) $\mathcal{H}_0$: *ComplexOnset, *ComplexCoda $\gg$ Max

(8) The Support

	*ComplexOnset	Max	*ComplexCoda
fɪag $\succ$ fag	L	W	

(9) $\mathcal{H}_1$: *ComplexCoda $\gg$ Max $\gg$ *ComplexOnset

The move from error (8) to new ranking (9) is done by a ranking algorithm, which for our current purposes could either be Biased Constraint Demotion (BCD, Prince & Tesar 2004) or Low-Faithfulness Constraint Demotion (Hayes 2004) or some amalgam of the two (see also Tessier 2007 and other references therein). As already stated, these algorithms are designed to keep all Markedness constraints above Faithfulness constraints to the extent possible given the observed data. In what follows, we gloss over the details of how these algorithms impose the persistent M $\gg$ F bias, but the above-cited works all provide such detail.

Our use of error-driven learning to understand how humans (rather than just simulated learners) acquire their phonology requires access to these errors, at one cognitive level or another. As encoded in (8), the algorithm needs to know the phonological content in both the observed winner form and the current grammar's loser form, as well as the violation profiles of each, which in turn derives its Ws and Ls. We do not believe, however, that the learning algorithm implies any child's conscious awareness of the difference between [fɪag] and [fag]: in other words, the learning procedure must know this much about errors, but as in many other linguistic domains this knowledge may be completely unconscious. The evidence in the acquisition literature on this matter is mixed. Some work provides explicit evidence that children can acknowledge their pronunciations as non-adult-like (e.g. Smith 2010, §7.3), which suggests they are at least aware of a difference between winners and losers in our technical sense, whereas other children show clear

resistance to such a distinction (e.g. the famous discussion of 'fish' vs. 'fiss' in Berko & Brown 1960). In the rest of this paper, we continue to use *the learner* to describe a learning procedure to approximate the relevant behavior of a particular child, Trevor, but we do not claim to know anything about whether Trevor knew what this learner was doing as he learned. We leave these very interesting questions as open ones for future research.[11]

To capture the data from §2, a BCD learner must be augmented with two additional properties. First, it must progress slowly and incrementally from one stage to the next, based on small re-rankings: when exposed to e.g. 'frog', 'toast' and 'string', it must not immediately acquire the variety of complex onsets and codas exemplified in those three words. Second, its individual stages must allow for the production of multiple outputs for the same input, mimicking the curves graphed in §2 — recalling our broad goal of producing this variation not in the grammar but in the learning mechanism, or rather in the way the learner uses the grammar. Our proposal to meet both of these goals is a version of Error-Selective learning (ESL, Tessier 2007, 2009) which uses an intermediate buffer called the Cache. The Cache creates incremental learning by restricting the speed and scope of the errors that are allowed to trigger the re-ranking algorithm. Below, we demonstrate how an ESL learner that uses its stored errors for grammar building can also change its outputs gradually, even when the grammars change categorically between $\mathcal{H}_n$ and $\mathcal{H}_{n+1}$.

3.1 How Error Selective Learning (ESL) Works

Every time the Error-Selective learner produces a new form with the current grammar (regardless of whether the form is target-appropriate or an error), the new form is stored in the Cache. Errors pile up until learning is triggered, which happens when some Markedness constraint has assigned Ls to some predetermined number of different Cached forms.

[11] We thank Yvan Rose and an anonymous reviewer for their challenges and questions about this topic.

(10) The Cache

	*ComplexOns	Dep	Max
fɹɑg ≻ fɑg	L		W
klɑk ≻ kæk	L		W
fɹɑg ≻ hʌwɑg	L	W	

If, in this sample Cache in (10), the predetermined threshold is set at 3, then the Cache will trigger learning via *ComplexOnset. Once triggered, the learner examines its Cache and selects an error (or small set of errors) to learn from. For this paper's purposes, this selection mechanism must simply choose an error that violates *ComplexOnset and as few other markedness constraints as possible. See Tessier (2009) for a more complete proposal about choosing errors to learn from.

Once chosen, this error or set of errors is added to the Support, which is the permanent repository of learning data:

(11) The Support

	*ComplexOns	Dep	Max
fɹɑg ≻ fɑg	L		W

The newly-updated Support in turn triggers the re-ranking algorithm to create a new ranking that will not make this particular error again.

An anonymous reviewer rightly wonders about the similarities and possible overlap between the Cache and Support on the one hand and the learner's lexicon on the other. We note that these stored errors do contain substantial amounts of lexical information, but neither can take over the lexicon's job. On the one hand, while the Cache will contain every surface form attempted by the learner, these errors must decay over time. On the other hand, while the Support's errors are permanently stored at least until learning is complete, they will only include a

fraction of encountered words (e.g., it cannot include any lexical item which never prompted an error.)

Returning to the example in (11): whatever the precise re-ranking effect, the Support is now slightly larger and the resulting grammar slightly more target-like – and with this new grammar $\mathcal{H}_{n+1}$, the learner begins to make new errors. At the same time as the Cache stores these novel forms, all of its stored forms made with the previous grammar $\mathcal{H}_n$ begin to decay. Each of these stored forms is associated with a relative probability, and these probabilities sink as a function of how many grammars ago each was produced (one may think of these probabilities as quantifying the strength or weight of each error). We propose that all forms enter the Cache at probability 1, and every time the learner produces an output it depresses the probabilities of all cached forms made with previous grammars by a consistent proportion. Note crucially that only old errors decay: errors made with the current grammar remain at probability (or weight) 1 until the next ranking is adopted.[12] As time passes, the probability of a cached form will be reduced to next to nothing, and it will be effectively forgotten. These properties of the Cache will be the key to our explanation of variability across stages.[13]

3.2 Gradual increase in faithfulness with ESL

Trevor's trajectory of faithfulness to obstruent-sonorant onset clusters was characterized in §2.1 as falling into three 'stages': (a) the *unfaithful* stage in which complex onsets are nearly completely banned, usually repaired by deletion; (b) the *variably increasing* stage across which faithful productions gradually appear and (c) the *faithful* stage at which (eventually) licit English complex onsets became uniformly preserved.

The unfaithful stage is a result of $\mathcal{H}_0$, which is determined by the M $\gg$ F bias of the ranking algorithm. Before any relevant errors have made it into the Support, this

[12] Among other things, this selective decay allows the ESL learner to learn effectively: when choosing an error to add to the Support, the learner will only examine those errors that have a probability of 1, being those that indicate problems the most current grammar has yet to resolve.

[13] This description of ESL differs from the description given in Tessier (2009) in at least two key ways: there, the Cache only stored errors, not all forms produced, and it was emptied after each cycle of re-ranking, while the Cache here changes more subtly.

bias ranks *ComplexOnset ≫ Max, meaning that all complex onsets are simplified and errors like /fɹag/ → [fag] begin to pile up in the Cache. As soon as an error like (8) has been added to the Support, the BCD algorithm will immediately build a grammar that produces complex onsets faithfully, as was shown in (7-9). Thus the question is how the *variably increasing* stage comes about in the interim.

We dub our approach the Dual Route ESL, because it derives the variably increasing stage from the two "routes" available to a learner for producing known words: run them through the current grammar to choose an output, or simply use a previously stored production from the Cache. At $\mathcal{H}_0$, the errors in the Cache all reflect the output of the unfaithful grammar, so both routes produce identical results. When $\mathcal{H}_1$ has been adopted, however, the Cache and the grammar are in conflict: the grammar protects complex onsets, but the majority of cached forms with complex onsets are previously-made errors with clusters reduced. As time passes, however, these errors will decay, and the grammar's faithful protection of onsets will win out.

A time slice in this process is illustrated in (12) below. Suppose at a time shortly after the adoption of $\mathcal{H}_1$, the learner attempts saying *frog*. If /fɹag/ has never previously been attempted it will not be cached in any way, so it will necessarily be parsed with $\mathcal{H}_1$, i.e. faithfully. If it does have a cached form, however, it will be an error created by $\mathcal{H}_0$, i.e. reduced to [fag].

(12) The dual-route ESL learner chooses between grammar and Cache

Learner wants to say "  "

40% chance: use the current grammar

/fɹag/	Dep	Max	*CO
a. ☞ fɹag			*
b. fag		*!	
c. fəɹag	*!		

60% chance: recycle from the Cache

	p	*CO	Dep	Max
fɹag ≻ fag	.6	L		W
…	…			
…	…			

If for example /fɹɑg/ has an entry in the Cache with p([fɑg]) = .6, then there is a 60% chance that this form will be recycled from the Cache and produced with a simplified cluster, and hence a 40% chance that its input will be run through the current grammar and produced with a faithful [fɹ] cluster.

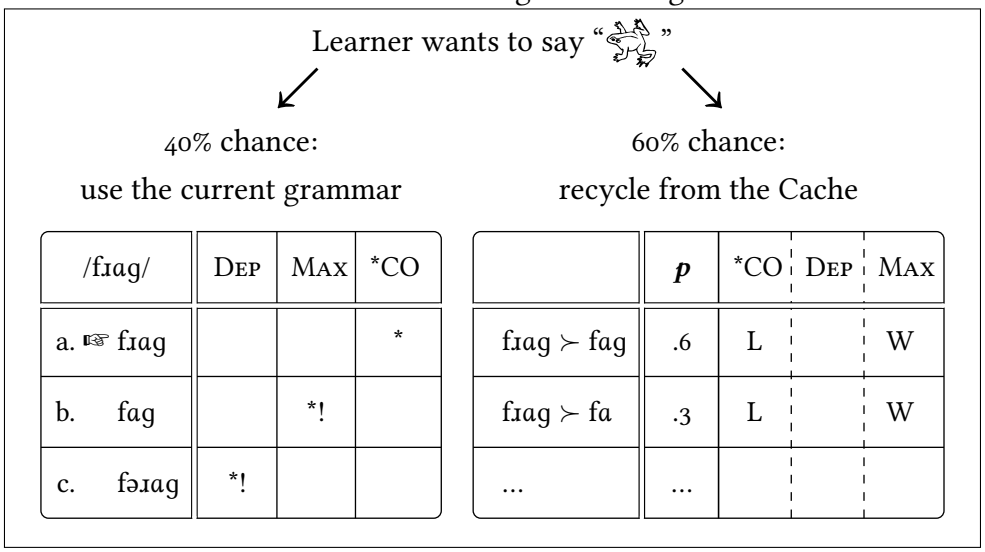
This snapshot reflects a moment during the variably increasing stage, where the grammar has reached the competence of the faithful stage, but the stored echoes of the unfaithful stage remain influential. As this example illustrates, the role of the Cache in production is to provide alternatives to the grammar – shortcuts, perhaps, to avoid using the grammar's latest machinery every time a word is attempted. With this in mind, we assume that those forms (errors or not) that are cached by the *current* grammar are not included in the calculation done in (12) above. Competition is therefore between decaying errors from previous grammars (those with probabilities of less than one) and the grammar's current choice of optimal outputs. Since the weights of old errors decay every time an input is attempted, the likelihood over time that a word will be produced via the Cached form route decreases, and so more often the current grammar's faithfulness will win out. Thus, the variably increasing stage eventually develops into the faithful stage, where Cached forms with reduced onset clusters have faded into negligibility and no longer interfere with the grammar's target ranking.

The Cache's sensitivity to individual words means that if a word has never been previously attempted, it will be produced with the current grammar. As two anonymous reviewers note, this prediction differs from that of a model in which variation is due to multiple stored grammars: The latter view does not distinguish between new and existing words. Trevor's corpus does not allow us to test our prediction, because we cannot assume that the first appearance of a word in our corpus represents Trevor's first use of it.

Before moving on, a somewhat technical point should be spelled out as to calculating the probability of using any cached error. Suppose that before the unfaithful stage of complex onsets illustrated above, the learner had an even earlier grammar which deleted coda consonants as well, meaning that *frog* was optimally produced as [fɑ] and cached as such. This error might still remain in the Cache at the time when the Max $\gg$ *ComplexOnset grammar is adopted to begin our

variably increasing stage, as in (13) below. Since [fɑ] is an older error it will have been decaying longer than the [fɑg], so its probability will be considerably lower (say, p([fɑ])=.3).

(13) The ESL learner stores errors from two grammars ago:



40% chance: use the current grammar

60% chance: recycle from the Cache

/fɹɑg/	Dep	Max	*CO
a. ☞ fɹɑg			*
b. fɑg		*!	
c. fəɹɑg	*!		

	p	*CO	Dep	Max
fɹɑg ≻ fɑg	.6	L		W
fɹɑg ≻ fɑ	.3	L		W
…	…			

A Cache like (13) gives the learner two older alternatives for uttering *frog*, with associated probabilities that are determined by their recency. The remaining probability mass is given over to the current grammar, and since the most recent error on *frog* is at .6, we give the grammar the remaining 40% shot at producing it. We do so by distributing the Cache's 60% chance among multiple cached errors for 'frog' according to their weights; in this case, [fɑg] has twice the probability of [fɑ] (.6 vs. .3), so [fɑg] gets 2/3 of the Cache's 60% chance.

The upshot of the Dual Route ESL is that changing the grammar will only gradually influence production, but over time the likelihood of using the current grammar steadily increases. Eventually, the next learning cycle is triggered, the grammar shifts to something slightly more target-like, and a gradual S-shaped progression towards faithfulness continues.

3.3 U-shaped development with ESL

The Dual Route ESL learner, as we have presented it, can account for increasingly faithful development along a particular dimension, but it has no way for productions to get increasingly unfaithful. Recall, however, that our stage-based description of Trevor's development along the KVT consonant harmony dimension consists of: (a) a *variably decreasing* stage, at which harmony is initially resisted but then increases in frequency, (b) a *variably increasing* stage at which harmony decreases, and then (c) a *faithful* stage at which harmony is gone. Recall also that for Trevor, KVT's harmony in fact occurs in two parts: first the process of K-dominant harmony emerges whereby *kiss* is produces as [kɪk], and then it is replaced by T-dominant harmony which creates [tɪs].

It may not be an accident that this special U-shape describes this pattern's development, since child consonant harmony is already special in a typological sense. Trevor's velar harmony assimilates Major Place features (Labial/Coronal vs. Dorsal) at a distance, across all vowels and some other consonants. While this is typical of child harmony in several languages, this kind of harmony is not attested in natural adult languages (Vihman 1978; Pater 1997b; Pater & Werle 2001; Hansson 2001). U-shaped development has also been discussed in the context of other child-specific harmonies (Kiparsky & Menn 1977; Menn 1983, see also Fikkert & Levelt 2008), velar fronting (Bleile & Tomblim 1991; Inkelas & Rose 2008), and other processes (Moskowitz 1970; Ferguson & Farwell 1975; Bernhardt & Stemberger 1998).

Our proposal for how child-specific processes are special is that they represent the effects of constraints induced by the learner, in response to the child's own productions and increased articulatory demands; see especially Inkelas & Rose's (2008) view of child-specific velar fronting; see also Goad (2001); Levelt & van Oostendorp (2007); Fikkert & Levelt (2008). We are thus claiming that Trevor was driven by concerns of some articulatory nature to induce the constraint Agree(KVT) somewhere around 1;5. As for the formal mechanisms the learner might have used, see Hayes (1999); Flack (2007); Hayes & Wilson (2008); Boersma & Pater (2007); Moreton (2010).

The questions of why this particular constraint was induced, and why at this moment, are ripe for speculation, but we are not sure that the nature of our data is sufficient to address them. While many children acquiring English adopt some consonant harmony patterns similar to Trevor's, it is by no means a given that they follow this trajectory — some show velar harmony earlier than other patterns rather than later, some show more labial harmony, and so on (see for example the differing patterns reported in Pater 1997a from Compton & Streeter 1977). Thus we suspect that the origins of AGREE(KVT) and its chronology will have to lie with something truly child-specific — e.g. the statistics of Trevor's own unique lexical input, or the raw frequency with which he attempted particular sequences of coronal and dorsal consonants in the weeks preceding the constraint's induction, the numerical subtleties of which this corpus cannot fully document. Given that at this age a child's productive lexicon is still rather small, it seems possible that the order in which Trevor and other children induce one or more AGREE constraints might come from individual idiosyncrasies, rather than any broader facts about English child-directed speech. We hope that future research will shed more light on this issue.

Regardless of its provenance, our account built from Trevor's observed data is that this constraint that drives consonant harmony is not present in the grammar at the beginning of the variably decreasing stage: at $\mathcal{H}_0$, shown in (14), there is no markedness pressure for [kɪs] to become [kɪk]. The learner generates faithful, disharmonic forms and stores them in the Cache, as in (15). These faithfulness constraints are essentially the same ones used by Pater & Werle (2003) in their analysis of Trevor's consonant harmony.

(14) $\mathcal{H}_0$: MAX(DOR), MAX(COR)

(15)

	MAX(DOR)	MAX(COR)
kɪs		

After a period of using the grammar and caching disharmonic forms, the learner decides that multiple consonantal places of articulation per word are too taxing

or otherwise intolerable, and so induces harmony constraints. For the KVT inputs under discussion here, we will call the induced harmony constraint AGREE(KVT), defined as in (16):

(16) AGREE(KVT): Assign one violation mark to a dorsal followed by a coronal[14]

How does the learner know how to rank their newly-induced constraint? We propose that adding a constraint to CON triggers a new application of the re-ranking algorithm using the existing Support. Since AGREE(KVT) has yet to be used by the grammar, there are no cached errors in which it has demonstrated its error-making capabilities, so the M ≫ F bias will ensure it is ranked at the top, with any other undominated Markedness constraints. The resulting $\mathcal{H}_1$ grammar will now drive harmony and start adding errors to the Cache:

(17) $\mathcal{H}_1$: AGREE(KVT) ≫ MAX(DOR) ≫ MAX(COR)

(18)

/kɪs/	AGREE(KVT)	MAX(DOR)	MAX(COR)
a. kɪs	*!		
b. tɪs		*!	
c. ☞ kɪk			*

We note that it is the initial K-dominance of the TVK words that prompts us to rank MAX(DOR) over MAX(COR) here. With this grammar update, the learner has entered the variably decreasing stage in earnest. At first, old errors have not yet decayed very much, so most of the time the cached route is chosen and harmony hardly ever applies. With increasing decay of cached forms, however, the grammar is increasingly used for production, causing errors like (18) to enter the Cache. As described in the previous section, these multiple forms are now in competition: the old faithful [kɪs], whose probability is decaying as time passes, and the new

[14] This AGREE constraint might be modified e.g. to apply only within syllables, or to ensure that adjacent segments in the clusters [kl] and [kɹ] are not affected. See also the somewhat different AGREE constraints used to capture Trevor's harmony in Pater & Werle (2001, 2003).

harmonic [kɪk], which the grammar will consistently produce on the increasingly-frequent occasions when it is asked to.

An anonymous reviewer notes that at this variably decreasing stage, cached forms are more adult-like than the form produced by the grammar; one might wonder why the child would ever prefer them. In our system, recycling errors from the Cache saves the learner from using the grammar and we conjecture that the "cost" of retrieving a word from the Cache is the same regardless of the form's complexity. Thus, an error may be retrieved and produced even if it violates a new and highly ranked markedness constraint.

Once the Cache has filled with harmonic forms, meaning enough errors with Ls assigned by AGREE(KVT) have accumulated in the Cache, they will trigger learning. An error like /kɪs/ → [kɪk] as in (18) above will be added to the Support (19).

(19)

	AGREE(KVT)	MAX(DOR)	MAX(COR)
kɪs ≻ kɪk	L		W

The resulting new grammar will rank AGREE(KVT) below MAX(COR) but not below MAX(DOR), so that T-dominant harmony will now be chosen as optimal:

(20) $\mathcal{H}_2$: MAX(COR) ≫ AGREE(KVT) ≫ MAX(DOR)

(21)

/kɪs/	MAX(COR)	AGREE(KVT)	MAX(DOR)
a. kɪs		*!	
b. ☞ tɪs			*
c. kɪk	*!		

As the K-dominant harmonic forms in the Cache decay, new T-dominant harmonic forms will be added. Once enough of these new errors accumulate, learning is

again triggered and the Support contains an error that prompts fully faithful [kɪs]:

(22) $\mathcal{H}_3$: Max(cor), Max(dor) ≫ Agree(KVT)

(23)

/kɪs/	Max(cor)	Max(dor)	Agree(KVT)
a. ☞ kɪs			*
b. tɪs		*!	
c. kɪk	*!		

Once this grammar is adopted, the learner is now in their last variably-increasing stage. Only the decaying cached errors prefer harmony, while the grammar has achieved adult-like competence, and thus faithfulness will increase steadily until harmonic forms disappear for good.

This section has augmented the dual-route ESL learner with a mechanism for producing regressions. However, an anonymous reviewer raises an important point: when an adult form violates multiple markedness constraints, demotion of one of them may cause what could be considered an increase in unfaithfulness; such cases do not require constraint induction, and are easily created by the basic ESL learner as presented in §3.2. For instance, the adult cluster [fr] may first be simplified to [f], with a faithful fricative; subsequent demotion of *ComplexOnset may cause the cluster to surface as [pr], with a stopped [p], as a result of a sonority sequencing pressure. In other words, the learner may become less faithful in some ways, while becoming more faithful in others. Such examples are particularly noted in Bernhardt & Stemberger (1998); Stemberger & Bernhardt (2001); Stemberger et al. (2001) and references therein. From our perspective, we stress that such cases do not require any active calculation by the learner of trade-offs between conflicting markedness or faithfulness, but rather fall out naturally from the workings of OT constraint conflict and the re-ranking algorithm.

3.4 KVT vs. TVK: Simultaneous U-shapes and S-shapes

Our treatment of Trevor's consonant harmony in §2.2 focused on KVT harmony, because the trajectory of this class of words significantly differed from an S-shape. Now that we have seen how this learner can produce a U-shape in faithfulness, we return to the comparison between KVT and TVK words made at the end of §2.2. What does our analysis of the U-shape predict for TVK words in the same system?

Since our view has been that each class of harmony targets (such as KVT words or TVK words) are targeted by a separate AGREE constraint, our account of the difference in U-shape vs. S-shape is rather straightforward. KVT words follow a U-shape because Trevor had not yet added AGREE(KVT) to CON at the onset of his recorded speech. TVK words do not show an initial resistance to harmony, therefore Trevor must have added AGREE(TVK) to CON before the recorded period. Since the content of the markedness constraints involved is rather narrowly stated, we cannot claim broader applicability beyond these specific consonant combinations, yet the outline of the analysis is generally applicable to any case of U-shaped development. We hope that future research will offer more insight as to the substantive content of these constraints.

That said, our constraint set does indeed make predictions about Trevor's development across stages, because the same faithfulness constraints, MAX(COR) and MAX(DOR), interact with both AGREE constraints. Recall from the two panels compared in (3) of §2.2 that for both KVT and TVK words, the switch from K-dominant to T-dominant harmony occurred roughly around 1;8. We illustrate now that this fact accords nicely with the constraint set we have used.

We start by noting that the initial grammar should now also include AGREE(TVK); this full $\mathcal{H}_0$ given in (24) produces the observed K-dominant harmony. Recall that it is the initial K-dominance of the TVK words that prompts us to rank MAX(DOR) over MAX(COR) here.

(24) $\mathcal{H}_0$: AGREE(TVK) $\gg$ MAX(DOR) $\gg$ MAX(COR)

When AGREE(KVT) is later added to the grammar, the full grammar has both markedness constraints at the top:

(25) $\mathcal{H}_1$: Agree(TVK), Agree(KVT) ≫ Max(dor) ≫ Max(cor)

The Cache in (26) below shows two representative errors made by this K-dominant $\mathcal{H}_1$, which we see taking over Trevor's productions around 1;6:

(26)

	Max(cor)	Max(dor)	Agree(KVT)	Agree(TVK)
kɪs ≻ kɪk	W		L	
dɔg ≻ gɔg	W			L

If either of these errors is added to the Support on some learning cycle, the constraint demotion algorithm will determine that Max(cor) must be ranked above one of the two Agree constraints — either KVT or TVK, depending on which error is selected. If the error [kɪk] from (26) is chosen, Agree(TVK) will remain undominated (as it is not violated by any version of 'kiss'), and beneath it will be ranked Max(cor). Since this ranking resolves the error [kɪk] in (26), the demotion algorithm is now free to install the remaining constraints according to its general M ≫ F bias, so that the new grammar will be as in (27):

(27) $\mathcal{H}_2$: Agree(TVK) ≫ Max(cor) ≫ Agree(KVT) ≫ Max(dor)

What is of interest here is that this new grammar will now produce T-dominant harmony for *both* classes of words. This is shown in (28) and (29) below:

(28)

/dɔg/	Agree(TVK)	Max(cor)	Agree(KVT)	Max(dor)
a. dɔg	*!			
b. gɔg		*!		
c. ☞ dɔd				*

(29)

/kɪs/	Agree(TVK)	Max(cor)	Agree(KVT)	Max(dor)
a. kɪs			*!	
b. kɪk		*!		
c. ☞ tɪs				*

This grammar is thus the one that emerges around 1;8. From this point on, T-dominant harmony will eventually generate more errors to learn from, leading both Agree constraints to be dominated by faithfulness, as in (30) and (31).[15]

(30)

/dɔg/	Max(cor)	Max(dor)	Agree(KVT)	Agree(TVK)
a. ☞ dɔg				*
b. gɔg	*!			
c. dɔd		*!		

(31)

/kɪs/	Max(cor)	Max(dor)	Agree(KVT)	Agree(TVK)
a. ☞ kɪs			*	
b. kɪk	*!			
c. tɪs		*!		

To summarize, we have shown an ESL-based OT analysis of Trevor's acquisition

[15]The progress that an ESL learner will follow, from the consistently T-dominant grammar in (27-29) to the fully faithful adult-like grammar in (30-31), contains a few further intermediate steps that we gloss over here. Since in ESL only a subset of the learner's cached errors are added to the Support, and since Max(cor) and Max(dor) are freely re-rankable with respect to each other (as the move from K-dominant to T-dominant forms requires), reaching $\mathcal{H}_3$ necessarily involves between one and three more learning cycles to reach the final harmony-less state. It is entirely possible that Trevor went through these additional stages, but that they are masked by the preponderance of stored forms he produces. We have limited our description of the data and the analysis to the steps that are sufficiently robust for statistical significance.

path, leading from initially-faithful KVT forms and K-dominant TVK forms into K-dominance throughout, followed by T-dominance throughout, and finally faithfulness. The U-shape acquisition of the KVT forms was caused by the introduction of AGREE(KVT) into the grammar, after an initial period with no markedness pressure to harmonize KVT words.

4 Simulation

Our analysis of Trevor's phonological development relies on two main theoretical tools: (a) the Cache, which stores previously uttered words, and creates smooth transitions in the child's production where there are discrete changes in the child's grammar, and (b) constraint induction, which introduces a markedness constraint into the grammar, and thus causes faithfulness to drop temporarily.

In this section, we offer a computational implementation of our analysis that takes Trevor's targets, and mimics his productions of these targets over time. The simulation Perl scripts and results are publicly available at `http://becker.phonologist.org/trevor/`. The implementation demonstrates that our theory is sufficiently detailed to be executed by a computer; at the same time, subtle mismatches between Trevor's behavior and the computer's behavior underline the complexity of the patterns involved and suggest directions for future research.

To demonstrate the basic S-shaped development, we start with a simulation of Trevor's complex onsets. The learner is given a list of Trevor's targets, an empty Cache, and a list of grammars to go through. In this case, there are two grammars: the M $\gg$ F grammar and the F $\gg$ M grammar.

The learner then proceeds to process the targets as explained in §3, starting by simplifying complex onsets and storing the outputs in the Cache. When the number of stored forms exceeds a threshold, the grammar is switched, at which point new words are produced faithfully. Older words, which are stored simplified in the Cache, start to decay, gradually giving way to the grammar.

The learner generates an output file that lists the targets that were attempted and

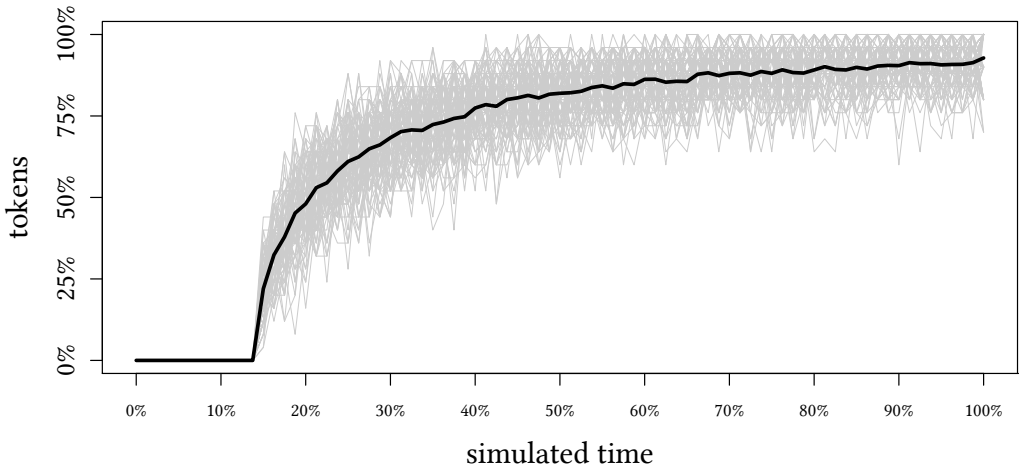


Figure 8: Simulation of faithfulness to complex onsets with ESL

the way they were produced, faithful or unfaithful. The learning was repeated 100 times, each time randomizing the order of the presentation of the tokens. Figure 8 shows the raw results of 100 simulations in gray. The black line traces the average rate of faithfulness for each day across all 100 simulations.

The results in Figure 8 show how a learner who switches sharply between grammars can still transition smoothly in production, replicating Trevor's behavior quite closely. We now turn to the second part of our analysis, where the introduction of a markedness constraint causes a U-shape dip in faithfulness.

To simulate the development of consonant harmony, the learner was given the list of Trevor's KVT and TVK target tokens, and four grammars to go through. The first grammar only contains one markedness constraint, AGREE(TVK), and the ranking MAX(DOR) ≫ MAX(COR), which causes K-dominant harmony on TVK words while allowing KVT words to surface faithfully. The second grammar introduces AGREE(KVT), causing K-dominant harmony throughout. In the third grammar, AGREE(KVT) and AGREE(TVK) are installed below MAX(COR), causing a shift to T-dominant harmony throughout. Finally, the AGREE constraints are installed below MAX(DOR) as well, and the grammar produces faithful outputs.

The learner went through Trevor's targets, deciding how to output each one given the grammar and the contents of the Cache, as described in §3. This simulation was repeated 100 times. The raw results are plotted in gray in Figure 9, with

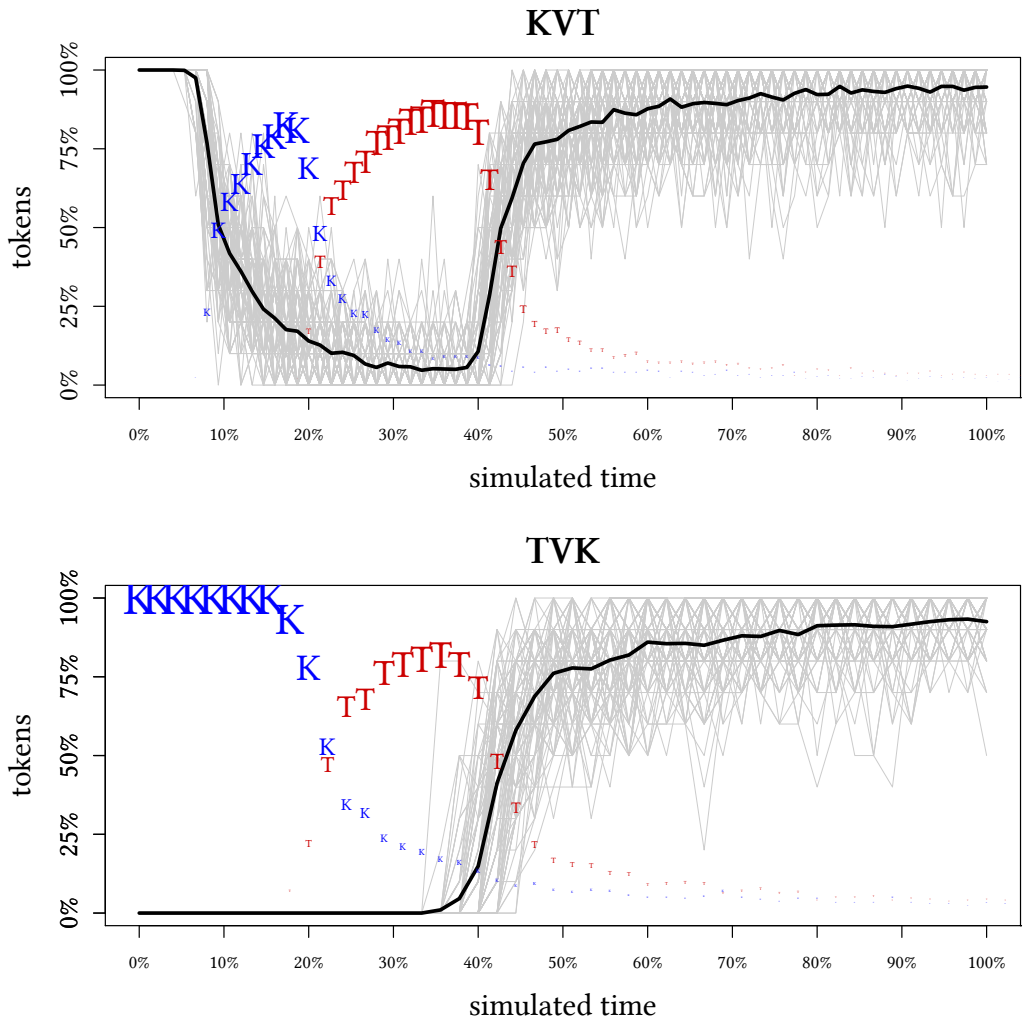


Figure 9: Simulation of resistance to consonant harmony with ESL

KVT targets separated from TVK targets. The black lines trace the average rate of faithfulness for each "day" across all 100 simulations, and the letter symbols indicate the average number of harmonized tokens on a given "day". We note that no particular statistical model is assumed here; the U-shape is a simple average of the simulations, and thus independent from the change point analyses and the associated approach to non-linearity we took in §2.4.

The results offer a satisfyingly close match to Trevor's behavior as seen in Figure 3 above: A clear U-shape appears in the KVT panel, and a clear S-shape appears in the TVK panel. The KVT panel begins with a faithful period, then a period of

K-dominant harmony, which smoothly transitions to a T-dominant harmony, and then into faithfulness again. The TVK panel begins with a K-dominant period, missing entirely any initial faithful productions. We note that the transition from K-dominance to T-dominance happens around the same time in both panels, also mimicking Trevor's behavior (discussed in §3.3 above).

We note a small qualitative difference between Trevor's behavior and the simulation: For Trevor, T-dominant harmony is rather weak among TVK words, and the T-dominant period is entirely subsumed by the K-dominant period, whereas in KVT words, T-dominant harmony is strong, and it only overlaps briefly with the K-dominant period. The simulated learner, however, makes the same transition in both panels, giving about equal overlap between K-dominance and T-dominance. Our simulated learner could implement Trevor's asymmetry by stipulating that K-harmonized TVK forms decay more slowly than others, thus sustaining their appearance in the output longer than otherwise warranted. Since we have no theoretical insight to offer on this point, we leave the learner in its pristine, symmetric form.

5 Conclusions

This paper has offered an analysis of Trevor's phonological development, showing that it includes processes which show S-shaped and U-shaped trajectories of faithfulness. We attributed S-shaped trajectories to the presence of markedness constraints in the grammar at the onset of speech, whereas U-shaped trajectories are caused by the introduction of a markedness constraint into the grammar after faithful forms are already produced.

These two kinds of developments are linked, in our view, to the typological status of these two patterns in adult languages, viz. that complex onset simplification is attested in adult languages but major place consonant harmony is not. In other words, an attested adult process is uniformly S-shaped in Trevor's speech, while a set of his child-specific processes show a much wider variety of developmental trajectories: An S-shape for TVK, PVK, and PVT words, a U-shape for KVT words and hardly any development in KVP and TVP words at all, appearing fully

faithful almost from the very start. In our analysis, AGREE(TVK), AGREE(PVK), and AGREE(PVT) already happen to be present in Trevor's grammar, and therefore we observe an S-shaped curve for those places. AGREE(KVT) is introduced late in the game, and thus we see its depressing effect on faithfulness in KVT words. For KVP and TVP words, we hardly see any unfaithful productions at all, which may be due to the lack of AGREE(KVP) and AGREE(TVP), or else may represent only the tail end of their U-shaped effect. We wish to point out this variety of developmental trajectories, and emphasize that introducing child-specific markedness constraints into the grammar in the middle of phonological acquisition can give rise to a U-shaped path. We find the connection between Trevor's child-specific place harmony and its U-shape to be highly suggestive, and likely not accidental; yet this one child's corpus, as rich as it may be, does not necessarily generalize to English and beyond.

Another central theme of this paper is the discrete, categorical grammar that gives rise to productions that are variable and gradual. The Cache, which was proposed in Tessier (2007, 2009) as a way to bring about realistic stages of acquisition into Constraint Demotion-based algorithms, finds here a new use as a buffer that smoothes over the categorical transitions between grammars (cf. the two lexicon approach, Menn 1983; Menn & Matthei 1992). This approach relegates variation to the learner's production mechanism; this is appropriate in the case of complex onsets and potentially-harmonic words of English, since the target adult grammar is of course categorically faithful. The question of how learners might acquire target grammars which are themselves variable remains open for future investigation, but we reiterate that a variable target grammar is different in kind from a variable developing grammar. It has been noted at least since Labov (1963) that adult grammars use variation to signal various sophisticated aspects of e.g. register and personal identity to their listeners. This purposeful *control* of variation, though often unconscious, is qualitatively different than the variability seen in a child's gradual mastery of English complex onsets and place of articulation. This distinction is reflected in our proposal: whatever the mechanisms for creating and constraining adult variation in the grammar turn out to be, the ESL learner has an extra-grammatical mechanism for creating variation along the way to mastery of their target.

Furthermore, since the Cache is designed to track the ranking arguments that come from individual lexical items, it creates variable behavior that is specific to lexical items, as is observed in Trevor's speech and many other known cases of variation in child or adult languages. Lexical items differ in their behavior, and yet the linguist can generalize meaningful patterns from these items using standard statistical methods, such as hierarchical regression models.

The computational simulation we offer shows that the grammatical apparatus we impute to Trevor generates the observed patterns, and allows rigorous testing of the theory. We add our results to a growing body of simulations of phonological development, noting that Trevor's corpus is particularly helpful in providing a single longitudinal record rather than cross-sectional experimental data (cf. Jarosz 2009) and that its richness in both types and tokens allows for a more direct account of lexical items rather than abstractions over e.g. syllable types or word shapes (cf. Boersma & Levelt 2000; Curtin & Zuraw 2001; Albright et al. 2008). Finally, we hope to have shown that a combination of quantitative analysis and a model of grammar can inform each other, and build a richer understanding of phonological development.

References

Albright, Adam, Giorgio Magri & Jennifer Michaels (2008). Modeling doubly marked lags with a split additive model. In *BUCLD 32: Proceedings of the 32nd annual Boston University Conference on Language Development*, Cascadilla Press. 36–47.

Baayen, R. Harald (2008). *Analyzing Linguistic Data: A practical introduction to statistics*. Cambridge University Press.

Bat-El, Outi (2010). Harmonic domains and synchronization in typically and atypically developing Hebrew-speaking children. Ms. Tel Aviv University.

Bates, Douglas & Martin Maechler (2009). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999375-32.

Berko, Jean & Roger Brown (1960). Psycholinguistic research methods. In P. Mussen (ed.) *Handbook of Research methods in Child Development*, New York: John Wiley. 517–557.

Bernhardt, Barbara H. & Joseph P. Stemberger (1998). *Handbook of Phonological Development: From the Perspective of constraint-based nonlinear phonology*. San Diego: Academic Press.

Bleile, Ken & J. B. Tomblim (1991). Regressions in the phonological development of two children. *Journal of Psycholinguistic Research* **20**. 483–499.

Boersma, Paul (1997). How we learn variation, optionality, and probability. In *IFA Proceedings*.

Boersma, Paul (1998). *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. Ph.D. dissertation, University of Amsterdam.

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 Eve V. Clark (ed.) *Proceedings of 30th Child Language Research Forum*, CSLI. 229–237.

Boersma, Paul & Joe Pater (2007). Constructing constraints from language data: The case of Canadian English diphthongs. Talk given at NELS38, University of Ottawa.

Bowerman, Melissa (1982). Starting to talk worse: Clues to language acquisition from children's late speech errors. In S. Strauss (ed.) *U-shaped behavioral growth*, Academic Press. 101–145.

Chiat, Shula (1983). Why mikey's right and my key's wrong: The significance of stress and word boundaries in a child's output system. *Cognition* **14**. 275–300.

Clahsen, Harold (1999). Lexical entries and rules of language: a multi-disciplinary study of German inflection. *Behavioral and Brain Sciences* **22**. 991–1013.

Compton, A. J. & Mary Streeter (1977). Child phonology: Data collection and preliminary analyses. In *Papers and Reports on Child Language Development 13*, Stanford, CA: Stanford University Press. 99–109.

Crawley, Michael J. (2007). *The R Book*. John Wiley & Sons.

Curtin, Suzanne & Kie Zuraw (2001). Explaining constraint demotion in a developing system. In Barbora Skarabela, Sarah Fish & Anna H.-J. Do (eds.) *Proceedings of BUCLD26*, Cascadilla Press. 118–129.

Dinnsen, Daniel & Mary Elbert (1984). On the relationship between phonology and learning. In Mary Elbert, Daniel Dinnsen & Gary Weismer (eds.) *Phonological theory and the misarticulating child*, Rockville, MD: ASHA Monographs, vol. 22. 59–68.

Ferguson, Charles A. & Carol Farwell (1975). Words and sounds in early language acquisition. *Language* **51**. 419–439.

Fikkert, Paula (1994). *On the Acquisition of Prosodic Structure*. Ph.D. dissertation, Leiden University.

Fikkert, Paula & Clara Levelt (2008). How does place fall into place? the lexicon and emergent constraints in children's developing phonological grammar. In Peter Avery, Elan Dresher & Keren Rice (eds.) *Contrast in phonology: Perception and Acquisition*, Berlin: Mouton de Gruyter. 231–268.

Flack, Kathryn (2007). *The sources of phonological markedness*. Ph.D. dissertation, University of Massachusetts Amherst.

Gnanadesikan, Amahlia (2004). Markedness and faithfulness constraints in child phonology. In René Kager, Wim Zonneveld & Joe Pater (eds.) *Fixing Priorities: Constraints in Phonological Acquisition*, Cambridge University Press. 245–291.

Goad, Heather (1997). Consonant harmony in child language: An optimality-theoretic account. In S.J. Hannahs & M. Young-Scholten (eds.) *Focus on phonological acquisition*, John Benjamins. 113–142.

Goad, Heather (2001). Assimilation phenomena and initial constraint ranking in early grammars. In Anna H.- J. Do, Laura Domínguez & Aimee Johansen (eds.) *Proceedings of the 25th Annual Boston University Conference on Language Development*, Cascadilla Press. 307-318.

Goldwater, Sharon & Mark Johnson (2003). Learning OT constraint rankings using a maximum entropy model. In Jennifer Spenader, Anders Eriksson & Osten Dahl (eds.) *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*. 111–120.

Hale, Mark & Charles Reiss (1998). Formal and empirical arguments concerning phonological acquisition. *Linguistic Inquiry* **29**. 656–683.

Hansson, Gunnar (2001). *Theoretical and Typological Issues in Consonant Harmony*. Ph.D. dissertation, UC Berkeley.

Hayes, Bruce (1999). Phonetically-driven phonology: The role of optimality theory and inductive grounding. In M. Darnell, E. Moravscik, M. Noonan, F. Newmeyer & K. Wheatly (eds.) *Functionalism and Formalism in Linguistics, Volume I: General Papers*, John Benjamins. 243–285.

Hayes, Bruce (2004). Phonological acquisition in optimality theory: the early stages. In René Kager, Joe Pater & Wim Zonneveld (eds.) *Constraints in Phonological Acquisition*, Cambridge: Cambridge University Press. 158–203.

Hayes, Bruce & Colin Wilson (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* **39**. 379–440.

Ingram, David (1974). Phonological rules in young children. *Journal of Child Language* **1**. 49–64.

Inkelas, Sharon & Yvan Rose (2008). Positional neutralization: A case study from child language. *Language* **83**. 707–736.

Jarosz, Gaja (2006). *Rich Lexicons and Restrictive Grammars - Maximum Likelihood Learning in Optimality Theory*. Ph.D. dissertation, Johns Hopkins University.

Jarosz, Gaja (2009). Effects of lexical frequency and phonotactics on learning of morphophonological alternations. Talk presented at Learning meets Acquisition: the learnability of linguistic frameworks from formal and cognitive perspectives, Osnabrueck, Germany.

Jaynes, E. T. (2003). *Probability Theory: The Logic Of Science*. Cambridge University Press.

Jesney, Karen (2007). The locus of variation in weighted constraint grammars. Poster given at the Workshop on Variation, Gradience and Frequency in Phonology, Stanford University.

Kazazis, Kostas (1969). Possible evidence for (near-)underlying forms in the speech of one child. In *Chicago Linguistics Society 5*. 382–388.

Kiparsky, Paul (2002). Syllables and moras in arabic. In Caroline Féry & Ruben van de Vijver (eds.) *The Syllable in Optimality Theory*, Cambridge University Press. 147–182.

Kiparsky, Paul & Lise Menn (1977). On the acquisition of phonology. In *Language Learning and Thought*, New York: Academic Press. 47–78.

Labov, William (1963). The social motivation of a sound change. *Word* **19**. 273–309.

Leopold, Werner F. (1939-1947). *Speech Development of a Bilingual Child*. Northwestern University Press.

Levelt, Clara (1994). *On the Acquisition of Place*. Ph.D. dissertation, Leiden University.

Levelt, Clara & Marc van Oostendorp (2007). Feature co-occurrence constraints in L1 acquisition. In Bettelou Los & Marjo van Koppen (eds.) *Linguistics in the Netherlands 2007*, John Benjamins. 162–172.

Macken, Marlys A. & Charles A. Ferguson (1981). Phonological Universals in Language Acquisition. *Annals of the New York Academy of Science* **379**. 110–129.

MacWhinney, Brian (1978). *The acquisition of morphophonology*, vol. 43. University of Chicago Press for the Society for Research in Child Development.

Maratsos, Michael (2000). More overregularizations after all : new data and discussion on Marcus, Pinker, Ullman, Hollander, Rosen and Xu. *Journal of Child Language* **27**. 183–212.

Marcus, G., S. Pinker, M. Ullman, M. Hollander, T. Rosen & F. Xu (1992). *Overregularisations in language acquisition*, vol. 57. Monographs of the society for research in child development.

Menn, Lise (1976). *Pattern, Control, and Contrast in Beginning Speech: A Case Study in the Development of Word Form and Word Function.* Ph.D. dissertation, University of Illinois, Urbana-Champaign.

Menn, Lise (1983). Development of articulatory, phonetic, and phonological capabilities. In Brian Butterworth (ed.) *Language production 2*, London: Academic Press. 3–50.

Menn, Lise & Edward Matthei (1992). The two lexicon model of child phonology: Looking back, looking ahead. In Charles A. Ferguson, Lise Menn & Carol Stoel-Gammon (eds.) *Phonological development :models, research, implications*, Parkton, MD: York Press. 211–247.

Moreton, Elliott (2004). Non-computable functions in optimality theory. In John J. McCarthy (ed.) *Optimality Theory in Phonology*, Blackwell. 141–163.

Moreton, Elliott (2010). Connecting paradigmatic and syntagmatic simplicity bias in phonotactic learning. Talk given at MIT, April 9.

Moskovitz, B. A. (1973). Acquisition of phonology and syntax. In J. M. E. Moravcsik K. J. J. Hintakka & P. Suppes (eds.) *Approaches to Natural Language*, Reidel Publishing Co. 48–84.

Moskowitz, Arlene I. (1970). The two-year-old stage in the acquisition of English phonology. *Language* **46**. 426–441.

Mudelsee, Manfred (2010). *Climate Time Series Analysis.* Springer.

Pater, Joe (1997a). *Consequences of Constraint Ranking*. Ph.D. dissertation, McGill University.

Pater, Joe (1997b). Minimal violation and phonological developement. *Language Acquisition* **6**. 201–253.

Pater, Joe & Jessica Barlow (2003). Constraint conflict in cluster reduction. *Journal of Child Language* **30**. 487–526.

Pater, Joe & Adam Werle (2001). Typology and variation in child consonant harmony. In Caroline Féry, Anthony D. Green & Ruben van der Vijver (eds.) *Proceedings of HILP 5*, University of Potsdam. 119–139.

Pater, Joe & Adam Werle (2003). Direction of assimilation in child consonant harmony. *The Canadian Journal of Linguistics* **48**. 385–408.

Prince, Alan (2002). Arguing optimality. ROA 562-1102.

Prince, Alan & Paul Smolensky (1993/2004). *Optimality Theory: Constraint Interaction in Generative Grammar*. Oxford: Blackwell. [ROA-537].

Prince, Alan & Bruce Tesar (2004). Learning phonotactic distributions. In René Kager, Wim Zonneveld & Joe Pater (eds.) *Fixing Priorities: Constraints in Phonological Acquisition*, Cambridge University Press. 245–291.

R Development Core Team (2007). R: A language and environment for statistical computing. Vienna, Austria. (http://www.R-project.org).

Rose, Yvan (2000). *Headedness and Prosodic Licensing in the L1 Acquisition of Phonology*. Ph.D. dissertation, McGill University.

Smith, Neil (2010). *Acquiring Phonology: A Cross-Generational Case-Study*. Cambridge University Press.

Smith, Neilson V. (1973). *The Acquisition of Phonology: A Case Study*. Cambridge University Press.

Smolensky, Paul & Géraldine Legendre (2006). *The Harmonic Mind: From Neural Computation To Optimality-Theoretic Grammar*. MIT Press.

Stemberger, Joseph & Barbara Bernhardt (2001). U-shaped learning in language acquisition, and restrictions on error correction. ROA 472.

Stemberger, Joseph, Barbara Bernhardt & Carolyn Johnson (2001). U-shaped learning in phonological development. ROA 471.

Steyvers, Mark & Scott Brown (2006). Prediction and change detection. In Y. Weiss, B. Scholkopf & J. Platt (eds.) *Advances in Neural Information Processing Systems*, MIT Press, vol. 18. 1281–1288.

Stoel-Gammon, Carol (1996). On the acquisition of velars in english. In *Proceedings of the UBC International Conference on Phonological Acquisition*, Cascadilla Press. 201–214.

Suomi, Kari, Juhani Toivanen & Riikka Ylitalo (2008). *Finnish Sound Structure*. Oulu University Press.

Tesar, Bruce & Paul Smolensky (2000). *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.

Tessier, Anne-Michelle (2007). *Biases and stages in phonological acquisition*. Ph.D. dissertation, University of Massachusetts Amherst, Amherst, MA.

Tessier, Anne-Michelle (2009). Frequency of violation and constraint-based phonological learning. *Lingua* **119**. 6–38.

Ullrich, Angela, Joseph Stemberger & B. May Bernhardt (2008). Variability in a german-speaking child as viewed from a constraint-based nonlinear phonology perspective. *Asia Pacific Journal of Speech, Language, and Hearing* **11**. 221–238.

Vihman, Marilyn (1978). Consonant harmony: Its scope and function in child language. In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.) *Universals of Human Language, Vol. 2: Phonology*, Palo Alto, CA: Stanford University Press. 281–334.