

Target selection in Error Selective Learning*

Michael Becker

michael@linguist.umass.edu

Abstract

SR (AKA 'faχaβ, Adam & Bat-El 2007a,b, 2009, 2010) shows a typical acquisition pattern of first *avoiding* marked structures, then *repairing* them, then producing them faithfully. I propose a method for diagnosing and quantifying avoidance, and then analyze the avoidance using the null parse approach. I offer a computationally implemented analysis of SR's acquisition path in terms of Error Selective Learning (Tessier 2007, 2008, 2009; Becker & Tessier 2011), modeling the avoidance of marked structure as selection of the null parse. The model predicts a realistic acquisition path using a persistent $M > F > M_{\text{PARSE}}$ bias and the relativization of M_{PARSE} to markedness constraints.

1 Introduction

Child speech is phonologically different from adult speech in several different ways. One prominent and noticeable property of child speech is deletion, typically leading to entire syllables missing in the child's pronunciation. It is a long-standing observation that by using deletion so liberally, children eliminate structures that they are not ready to pronounce, such as codas, complex onsets, and unstressed syllables. As they approximate the adult pronunciation, children gradually phase deletion out of their speech. Another option children have, however, is to avoid words that have such structures altogether; they stay silent, or choose to say different words.

The idea that avoidance is a way to approximate adult speech is expressed in Kiparsky & Menn (1977:56–58), quoted by Kager et al. (2004:11), emphasis mine:

- (1) Different children exclude **definable classes of output by different means** ... When we observe such repeated 'exclusion', we conclude that these classes of outputs ... represent difficulties to the child, and the various rules of child phonology (substitutions, deletions, etc.) as well as **selective**

*Thanks to Outi Bat-El for sharing her data and her insights, and to two anonymous reviewers for their helpful comments. Thanks also to Adam Albright, Hannes Fellner, Wendell Kimper, Joe Pater, John McCarthy, and Matt Wolf for feedback and valuable comments. Special thanks go to Anne-Michelle Tessier for thorough discussion of the proposal. I also thank the audience at the workshop on computational modeling of sound pattern acquisition (University of Alberta, Edmonton), and at the UC Berkeley linguistics department, in particular, Bruce Hayes, Sharon Inkelas, and John Sylak. There are no remaining errors.

avoidance of some adult words, are devices the child finds for dealing with those difficulties.

The quote is particularly insightful given that it predates a theory with output constraints that can be satisfied in a variety of ways; the idea of excluding classes of output by different means was not expressible in the rule-based theory that was prevalent at the time (deletion or avoidance could be modeled separately in rule-based theory, but the conspiracy between them could not). In terms of Optimality Theory (Prince & Smolensky 1993/2004), a single markedness constraint, say against complex onsets, causes both simplification of these complex onsets by deletion, and avoidance of words that have complex onsets in their adult form. By avoiding words with complex onsets in their adult forms, children eliminate the faithfulness violations that would be caused by simplifying the adult form.

Target avoidance, or target selection, is mentioned as a feature of language acquisition in, among others, Ferguson & Farwell (1975); Kiparsky & Menn (1977); Schwartz & Leonard (1982); Schwartz (1988); Schwartz et al. (1987); Stoel-Gammon & Cooper (1984); Fikkert (1994); Grijzenhout & Joppen-Hellwig (2002); Menn (2004); Goad & Rose (2004). In particular, Adam & Bat-El (2007a,b, 2009, 2010) discuss SR's target avoidance in the context of prosody (preferences for trochaic targets) and selection of vowels (preference for [a]).

By selecting targets based on phonological criteria, the child carves out an increasingly representative subset of the adult language for their productions. In SR's case, he starts with a lexicon that has very few final sonorants in its adult forms, which is not at all representative of adult Hebrew. After a period of avoiding these words, he gradually adds final sonorants to his speech, until such words make more than a third of all the words he says, as in the adult language.

In this paper, I focus on two structures in SR's speech, initial complex onsets and final sonorant codas. I show that he avoids words that have these structures in §2, and I offer an Optimality Theoretic analysis that derives the avoidance by selection of a null parse, or a failure to parse the input, in §3. A generalized theory of avoidance, with a learning algorithm that produces the stages of acquisition, from avoidance to deletion to adult speech, is presented in §4. A computational implementation of the learning algorithm, which formalizes the analysis, is presented in §5, and §6 concludes.

A particularly interesting aspect of SR's development is that complex onsets are avoided longer than sonorant codas. Theoretically, this will require avoidance to be sensitive to the marked structure that is being avoided. The model will accommodate this fact by relativizing a constraint against avoidance to particular markedness constraints that are active in the grammar.

2 Diagnosing target selection

By its very nature, target selection is not directly detectable. Rather, it needs to be deduced from the data by identifying the conspicuous underrepresentation of certain types of words. The phonological acquisition literature is mostly interested in unfaithful productions, and rightly so, since they are directly observable.

As the child transitions from avoiding a certain structure to a more advanced stage (either simplified production or adult-like production), the previously underrepresented structure will become increasingly prevalent, and gradually reach its ratio in the adult language. On the other hand, structures that are not avoided will be present in the data from the beginning, and their relative prevalence will not increase over time (in fact, it may decrease over time, as previously avoided forms become more prevalent).

SR's longitudinal corpus is ideally suited for testing the relative proportion of various structures along the acquisition path. SR was recorded by his aunt about once a week, starting from age 1;2. His productions were transcribed and paired with their intended meaning (as determined by his aunt). Most words were produced in the context of directed play, including naming tasks.

Figure 1 charts the development of SR's initial complex onsets. The diamonds trace the production of complex onsets as a percentage of SR's total productions. Each diamond represents one session, with its size proportional to the number of words recorded on that session. Up until the age of 1;7, SR does not produce complex onsets at all. Once he starts producing them faithfully, they gradually increase in frequency until they make about 2% of his productions. Statistically speaking, this development of SR's productions can be shown to be a positive correlation between SR's age and the probability that he would produce a complex onset. A logistic regression model using the *glm* function in R (R Development Core Team 2011) confirms that the correlation is significant ($z = 6.89, p < .0001$).

The circles in Figure 1 trace SR's attempts to produce words with complex onsets. If SR were to choose words from the adult language based purely on their meaning, we would expect to see no progression; words with complex onsets would be attempted at a constant rate throughout the period of acquisition. This is not the case: SR starts by hardly attempting any complex onsets at all until about 1;5, then his attempts increase gradually until about 1;8. This gradual rise, or positive correlation between age and attempts of complex onsets, is statistically significant. A logistic regression model with a change point¹ at 1;8 shows an initial positive

¹Change points allow a regression analysis to model the rise and subsequent fall in the ratio of complex onsets in the data. In the case of SR's complex onsets, the introduction of the change point is justified by the significant improvement it makes over the model that lacks a change point

slope from 1;2 up to 1;8 that is highly significant ($z = 3.84, p < .0005$). Admittedly, the following decline in the prevalence of complex onsets is left unexplained here. I speculate that the decline from 1;8 to 2;2 is due to the acquisition of other marked structures, or perhaps other morpho-phonological categories (such as verbs), which make complex onsets relatively less common. I leave this puzzle for future work.

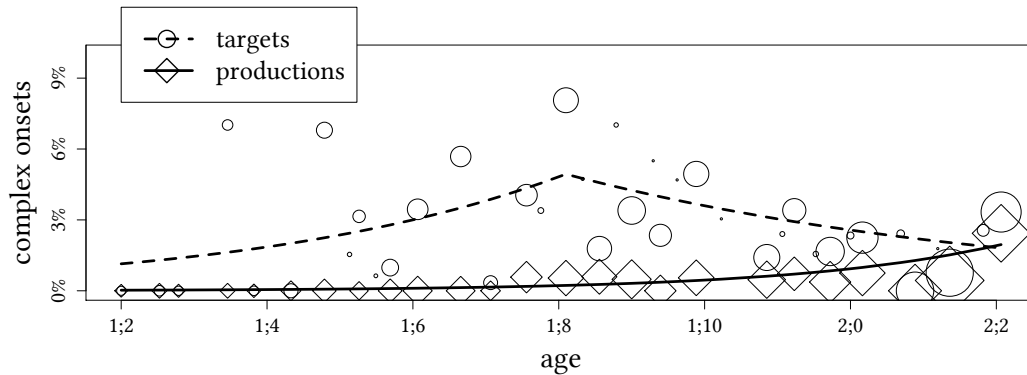


Figure 1: Increase in SR's initial cluster targets up until 1;8

To summarize SR's treatment of complex onsets, we have seen that he starts without any complex onsets, neither in production nor in attempts. In other words, it is not simply the case that words with complex onsets are simplified — they are eliminated altogether, or avoided. The initial state gradually gives way to a stage where words with complex onsets are attempted, and produced unfaithfully. The increase in attempts of complex onsets, or the decrease in their avoidance, is statistically significant.

SR's treatment of final sonorant codas shows a similar pattern of avoidance. Figure 2 charts the development of his final sonorants, again showing productions with a solid line and attempts with a dashed line. Until the age of 1;4, SR hardly produces any sonorant codas at all, and then he quickly transitions to producing them mostly faithfully at 1;6. This acquisition is much quicker than SR's acquisition of complex onsets, which is to be expected given the higher frequency of these sonorant codas. The increase in final sonorants is also seen in his attempts: words with final sonorants in their adult form become increasingly prevalent in his speech, up until 1;8, and again this increase comes out statistically significant in a logistic regression model ($z = 6.68, p < .0001$).

The evidence for avoidance of final sonorants is precisely the increase in SR's

($\chi^2(1) = 30.2, p < .0001$). For a similar use of a change point analysis in child data, see Becker & Tessier (2011). For general reference on regressions with a change point, see Baayen (2008) §6.4, Jaynes (2003) §15.8.1, and Mudelsee (2010) §4.2.2.

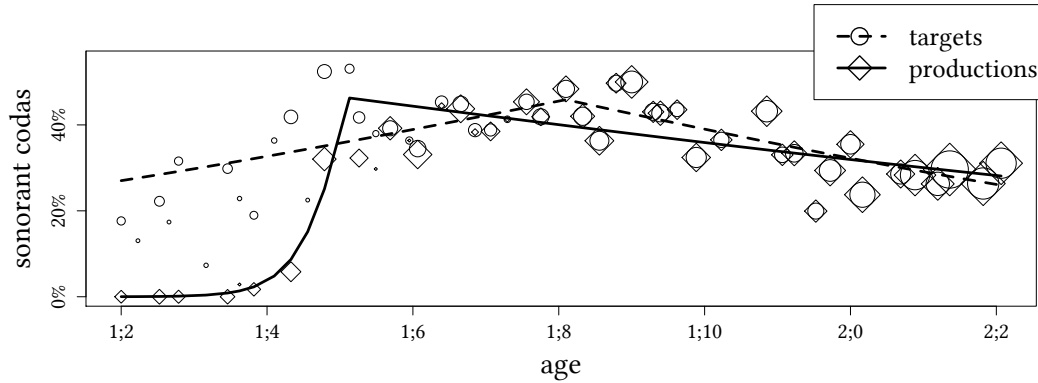


Figure 2: Increase in SR's final sonorant targets up until 1;8

attempts from 1;2 to 1;8. If SR had not avoided sonorant codas, they would be targeted or attempted at a constant rate, but the regression analysis confirms that the change in the ratio of sonorant codas is significantly different from zero. Past the age of 1;8, the ratio of sonorant codas declines both in attempts and in production — mirroring the decline seen for the complex onsets.

The evidence for avoidance relies exclusively on the distribution of attempted targets, i.e. on the adult phonological forms, disregarding SR's surface productions. How SR pronounced these words is orthogonal to the question of avoidance; the fact that they are conspicuously missing from his early months of speech is enough evidence for target selection. Equivalently, the question can be framed in terms of overrepresentation of the complement forms: a vocabulary limited to final vowels and obstruents, with no initial clusters, is not a representative sample of the Hebrew lexicon. Now, it is possible that such a vocabulary is representative of a *child* lexicon, and that this is an accident about the Hebrew lexicon that concepts that children discuss at that age are phonologically simple. This does not seem to be the case: words like 'apple' and 'pear' (which SR says) are hardly more relevant to a child's interests than 'water' and 'ice-cream' (which SR avoids). A cross-linguistic study of the acquisition of these concepts would clear this point, but this is beyond the scope of this paper.

SR does not attempt all words with equal probability. Rather, he carves out a non-representative portion of the lexicon for his early productions, limiting his attempts to words that he can produce with sufficient accuracy. To be sure, this is not a claim about SR's conscious efforts; as with most phonological behavior, it is entirely plausible that his avoidance is unconscious. The analysis provided below does not assume any conscious manipulation of linguistic elements.

While the evidence for avoidance comes from the distribution of SR's targets, it is also instructive to look at his productions. Once words with final sonorants

or initial clusters appear, what form do they take? Here, we have to rely on the transcription of SR's words. While it is known that adults are imperfect transcribers of child productions, we have no better data for SR's speech, and therefore I will assume that the transcriptions are at least indicative of SR's actual speech.

Consonant deletion is SR's most common repair for word-final sonorants and word-initial clusters. For final sonorants, deletion accounts for about half of all unfaithful tokens, as in (2).² With initial onsets, deletion accounts for about three quarters of all unfaithful tokens, as exemplified in (3).

(2) SR's repairs of final sonorants by deletion

Adult target	SR's production		
'pil	'pi	1;04.10	'elephant'
na.'meɸ	na.'me	1;05.08	'tiger'
'ma.im	'ma.i	1;05.15	'water'

(3) SR's repairs of initial clusters by deletion

Adult target	SR's production		
'pkak	'pak	1;05.29	'cap'
'gli.da	'gi.da	1;06.12	'ice-cream'
dvo.'ɸa	do.'ɸa	1;06.26	'bee'

If we accept that consonant deletion is used as a repair for both final sonorants and initial clusters, we must conclude that the same faithfulness constraint, namely MAX-C, is involved in both processes. The analytical implications of this shared reliance on MAX-C are explored in §3.2. Going beyond the specifics of SR's particular acquisition path, however, it seems highly likely that a single repair, and in particular deletion, will be deployed by a child to deal with two separate sources of markedness; thus, the problem at hand is likely to be relevant to the analysis of acquisition paths in general. Even if the reader does not accept the accuracy of the transcriptions in this corpus, the form of the problem remains if one accepts that deletion can be used to repair more than one marked structure.

3 Analysis: SR's avoidance as a null parse

I will analyze avoidance in OT (Prince & Smolensky 1993/2004), using the idea that avoidance results from inputs of certain kinds being mapped onto the null parse (Prince & Smolensky 1993/2004; McCarthy & Wolf 2010). The null parse

²SR opts for metathesis in about one third of the tokens when it repairs an onsetless unstressed syllable, as in [ʃa.li] or [ʃa.li] for adult [ʃa.il] 'Lyle'.

is a candidate that fails to have any phonological or morphological structure, and thus vacuously satisfied all markedness and all faithfulness constraints. Noted \odot , the null parse only violates the special constraint M_{PARSE} .

Prince & Smolensky (1993/2004) proposed the null parse to model phonotactic gaps, and the approach has been used this way by Smith (2009) and Albright (2011). The null parse has also been used to model paradigm gaps, e.g. in Rice 2006, Bat-El 2010 (though see Albright 2010 for a different view). The null parse formalizes the idea that a phonological derivation may not succeed in generating an output; passing a phonological or morpho-phonological input form through the grammar produces no output, or equivalently, produces an output with no phonological or morphological form at all.

3.1 Error-selective learning of target avoidance

To generate realistic acquisition paths based on the null parse, I use Error Selective Learning (Tessier 2007, 2008, 2009; Becker & Tessier 2011), which is an elaboration of the error-based learning algorithms in Prince & Tesar (2004), Hayes (2004), et seq. In error-based learning, an error is defined as a mismatch between an adult form and the pronunciation is it assigned by the child. The child uses the error to learn by pairing it with the adult form, making a winner-loser pair, and finding out which constraints prefer the winner using comparative tableaux (Prince 2002). The winner-loser pairs are stored in the Support, which represents the evidence that the child has for their current constraint ranking. What Error Selective Learning adds to the basic algorithms in Prince & Tesar (2004) and Hayes (2004) is a method for choosing errors to learn from in a gradual way. In this paper, I will simply assume that the errors are selected correctly; the interested reader should consult the works cited above for further detail. One property of Error Selective Learning is that more frequent errors trigger learning earlier. In the case of SR, Error Selective Learning ensures that learning matches SR's path of mastering sonorant codas first and complex onsets later, simply because sonorant codas are more frequent in the target language.

To see how Error Selective Learning works, I start with a simplified grammar that has only three constraints: the markedness constraint M , the faithfulness constraint F , and M_{PARSE} . There will only be one word in the language, 'adult form', which violates M .

The learner is equipped with a $M > F > M_{\text{PARSE}}$ bias, meaning that in the absence of evidence to the contrary, all markedness constraints are at the top, all faithfulness constraints are ranked below, and M_{PARSE} is at the bottom. With M_{PARSE} so lowly ranked, the null parse is always the winner, and all words are avoided. This initial

grammar, also known as the initial state, is noted as \mathcal{H}_0 in (4). With this grammar, the child is silent, as all words are mapped onto the null parse.

(4) $\mathcal{H}_0: M \gg F \gg \text{MPARSE}$

MPARSE is at the bottom, everything is avoided.

When the child attempts to pronounce ‘adult form’, the current grammar selects the null parse as the winner (5). The outside observer hears nothing being said, so the word ‘adult form’ is avoided. Since we assume ‘adult form’ as the input, the candidate ‘adult form’ is faithful, so it only violates M, but not F. By definition, the null parse only violates MPARSE.

(5)

/adult form/	M	F	MPARSE
a. adult form	*!		
b. \emptyset			*

The adult form does not win in (5), i.e. the child makes an error. To learn from this error, the intended winner is compared to the actual winner, and this comparison creates the comparative tableau in (6), also known as the Support. The Support collects the errors that the child decides to learn from. The constraint M assigns a violation mark to the winner but not to the loser, so it prefers the loser. This preference is marked with an L in (6); similarly, MPARSE prefers the winner, so it marked with W. When the ranking algorithm is run on the Support, MPARSE is installed, meaning it is placed at the top of the hierarchy. The single winner-loser pair is accounted for and thus removed, leaving M and F without any evidence about their ranking. Due to the $M > F$ bias, M is installed first below MPARSE, and F is installed below M, giving the new grammar in (7).

(6)

	M	F	MPARSE
adult form $>$ \emptyset	L		W

(7) $\mathcal{H}_1: \text{MPARSE} \gg M \gg F$

Faithfulness at the bottom, everything is simplified

In this new grammar, faithfulness is lowly ranked, so running ‘adult form’ through the new grammar will give a new winner ‘simplified form’. We will assume that ‘simplified form’ does not violate M and does violate F. Choosing ‘simplified form’ as the output is still an error, so learning needs to take place again. Now the learner pairs ‘simplified form’ with the adult form, and learns from the new error. The winner-loser pair from (6) is still part of the Support, so it also appears in (8).

(8)

	M	F	MPARSE
adult form > \odot	L		W
adult form > simplified form	L	W	

When running the ranking algorithm on (8), we have two constraints that only prefer winners to choose from, F and MPARSE. Due to the $F > \text{MPARSE}$ bias, we first install F, followed by MPARSE. When both are installed, M is free for installation below them, as in (9). At this point, the learner reached the adult grammar.

(9) \mathcal{H}_2 : $F \gg \text{MPARSE} \gg M$
 Adult forms surface faithfully

The learning path as illustrated moved the learner from avoidance through simplification to the adult form. The three stages appear without anything known about the specific markedness and faithfulness constraints involved; the trajectory of the learning path is only due to the general definition of markedness and faithfulness and to the $M > F > \text{MPARSE}$ bias in the constraint ranking algorithm.

The ranking algorithm as given by Prince & Tesar (2004) and Hayes (2004) only use an $M > F$ bias, which I augmented with an $F > \text{MPARSE}$ bias. The addition of further biases to constraint ranking algorithms has several precedents, such as the addition of the $\text{OO-F} > M$ bias and the splitting of F into a specific $F > \text{general F}$ bias; see Tessier (2006, 2007) for a review.³

The null parse approach has only been used to model avoided forms in adult language, to my knowledge. My use of it here is particularly close to its use in the modeling of phonotactic gaps, as in Prince & Smolensky (1993/2004), Smith (2009) and Albright (2011). This paper integrates the null parse approach with the learning algorithms used in constraint-based approaches.

The idea that gaps may shape language acquisition suggests that over time, some languages will get rid of words that have marked structures. An instructive case comes from the history of rhotic-initial roots in Proto-Indo-European: in most daughter languages, these rhotic-initial roots appear with prothesis, but the same roots are missing in Anatolian (Rasmussen 1999; Operstein to appear). Perhaps future attention from historical linguists will uncover more of these cases.

³The proposed $F > \text{MPARSE}$ bias is motivated in this case by SR's acquisition path, and any other acquisition path that is similarly characterized by avoidance followed by simplification. It is hoped that this bias will receive additional support in the future from other sources, cross-linguistic and experimental.

3.2 Application to syllable margins

In the previous section, we have seen how a single markedness constraint can interact with faithfulness and MPARSE to produce the kind of learning trajectory that SR exhibits, with avoidance followed by repair and then mastery.

This section assumes further that SR's initial clusters and final sonorants are both repaired with deletion, and thus both give rise to violations of the same faithfulness constraint, MAX-C. Admittedly, it may not be possible to verify that deletion is involved in both cases beyond reasonable doubt, but the general case of two different markedness violations giving rise to violation of the same faithfulness constraint seems quite plausible. In particular, deletion is known to be highly prevalent in child language, and it is quite likely that at least some children deploy deletion as a repair for more than one marked structure.

With sonorants in word-final codas, SR's main choices are shown as candidates in (10). The grammar is in its initial state, as in (4), with $M \gg F \gg \text{MPARSE}$. Here, the markedness constraints are instantiated by *SONCODA, which penalizes final sonorants, and *COMPLEX, which penalizes initial clusters. The faithfulness constraint is MAX-C, which assigns one violation mark for every consonant of the input that lacks an output correspondent. The first candidate in (10) is fully faithful, violating the markedness constraint *SONCODA. The second candidate is a simplified form, with the final consonant deleted. The final candidate is the null parse, which only violates MPARSE.

(10)

/na.'meɸ/	*SONCODA	*COMPLEX	MAX-C	MPARSE
a. na.'meɸ	*!			
b. na.'me			*!	
c. ☐				*

When SR moves to the next grammar, as in (7), MPARSE is installed above *SONCODA. *COMPLEX is not involved, and it stays at the top of the hierarchy. But since MPARSE is no longer at the bottom, complex onsets can no longer be avoided; this unintended consequence is shown in (11).

(11)

/'gli.da/	*COMPLEX	MPARSE	*SONCODA	MAX-C
a. 'gli.da	*!			
b. ☐ 'gi.da				*
c. ☐		*!		

Once MAX-C is below MPARSE, avoidance is more costly than deletion, and all

marked structures that can be repaired by deletion can no longer be avoided. In other words, once sonorant codas are simplified, complex onsets should be simplified as well. Yes this is not the case: SR starts simplifying sonorant codas from the very first recording session at 1;2, while simplified complex codas don't appear until more than a month later.

The unintended interaction between the processes seen above is more severe when examined in a more generalized way. In §4 below, I show the true depth of the problem, and offer a solution in the form of sub-categorization of MPARSE.

4 Generalizing the MPARSE(M) approach

In this section, I show that the $M > F > \text{MPARSE}$ bias is too coarse to derive a realistic acquisition path; the effect of MPARSE is too dramatic. Instead, I will propose that MPARSE is relativized to particular markedness constraints, thus limiting its effect.

4.1 The problem: avoidance lost too soon

To see the problem in its general form, assume a more realistic grammar with several markedness constraints, noted M_1, M_2, M_3 , etc., and several faithfulness constraints, noted F_1, F_2, F_3 , etc.

In the initial state, as in (12), all the markedness constraints are in one stratum at the top, followed by all of the faithfulness constraints, and MPARSE at the bottom.

$$(12) \mathcal{H}_0: \quad M_1, M_2, M_3, \dots \gg F_1, F_2, F_3, \dots \gg \text{MPARSE}$$

Initial state: everything avoided.

Now, the learner notices that one word, 'adult form', is being avoided, and it is paired with the current output, the null parse, to form the winner-loser pair in (13). We assume that 'adult form' violates M_1 . The markedness constraints that don't have an L in their column are installed first, followed by MPARSE. Once the winner-loser pair is accounted for, M_1 is free to be installed, followed by all of the faithfulness constraints. The resulting grammar is in (14).

In this new grammar, MPARSE outranks all of the faithfulness constraints, so the null parse will never be a winner. Any possible repair for M_1 will be preferred to a violation of MPARSE. What happened was that learning triggered by a single markedness constraint caused avoidance to no longer be available as a response to any markedness violations.

(13)

	M ₁	M ₂	M ₃	F ₁	F ₂	F ₃	MPARSE
adult > ∅	L						W

- (14) $\mathcal{H}_1: M_2, M_3, \dots \gg \text{MPARSE} \gg M_1 \gg F_1, F_2, F_3, \dots$
 Nothing avoided; everything simplified.

Continuing from here, the learner now produces ‘adult form’ simplified. Pairing ‘simplified form’ with ‘adult form’ adds another winner-loser pair to the Support, as in (15).

(15)

	M ₁	M ₂	M ₃	F ₁	F ₂	F ₃	MPARSE
adult > ∅	L						W
adult > simplified	L			W			

When running the ranking algorithm on (15), M₂ and M₃ are free for installation first. Then F₁ and MPARSE need to be installed to free M₁, followed by the remaining faithfulness constraints F₂ and F₃. The resulting grammar is in (16).

- (16) $\mathcal{H}_2: M_2, M_3, \dots \gg F_1 \gg \text{MPARSE} \gg M_1 \gg F_2, F_3, \dots$
 M₁-violating forms surface faithfully; everything else simplified if F₁ is not involved; avoided if F₁ is involved.

In this grammar, the low ranking of M₁ allows forms that violate it to surface faithfully. Avoidance in general is not in use, because most faithfulness constraints are below MPARSE, but avoidance can happen in a situation where the only other option is violating F₁.

The learning path described in (12-14-16) seems rather implausible. While it correctly starts with avoidance, and leads particular forms through simplification to the adult forms, there is an unintended consequence for avoidance in general. The early promotion of MPARSE above all faithfulness constraints means that avoidance is no longer available as a strategy for dealing with any kind of phonotactic restriction. Since the null parse will no longer be a winner once the grammar in (14) is adopted, there will no longer be evidence for ranking it below any markedness constraint other than M₁.

4.2 The solution: Markedness-based avoidance

We have seen in §4.1 that MPARSE must not be promoted too early, or it will have an implausibly strong effect. The solution I propose relies on a better formalization

of Kiparsky & Menn’s insight quoted in (1), according to which avoidance is a response to a markedness constraint.

Instead of installing MPARSE above markedness as in §4.1, the revised algorithm installs a markedness-specific instances of MPARSE. This approach finds precedent in Rice (2006), who also proposes that MPARSE should be relativized to apply in fewer contexts.

The learner starts as before, with a grammar that has several markedness constraints, then several faithfulness constraints, and finally MPARSE at the bottom of the hierarchy.

- (17) \mathcal{H}_0 : $M_1, M_2, M_3, \dots \gg F_1, F_2, F_3, \dots \gg \text{MPARSE}$
 Initial state: everything avoided.

Again the learner’s grammar generates the null parse from the input ‘adult form’, and a winner-loser is formed, as in (18).

(18)

	M ₁	M ₂	M ₃	F ₁	F ₂	F ₃	MPARSE
adult > ∅	L						W

Now, applying the ranking algorithm would normally lead us to install MPARSE. But instead, we create a new constraint, MPARSE(M₁), and install it. This new constraint only penalizes the null parse if the fully faithful candidate violates M₁ (see 23 below for a formal definition). This frees up M₁, and following it, faithfulness can be installed, followed by the general MPARSE. With this new grammar, faithfulness still outranks general MPARSE, and avoidance is still largely in place. Only forms whose fully faithful candidate violates M₁ are simplified.

- (19) \mathcal{H}_1 : $M_2, M_3, \dots \gg \text{MPARSE}(M_1) \gg M_1 \gg F_1, F_2, F_3, \dots \gg \text{MPARSE}$
 M₁-violating adult forms simplified; everything else avoided.

Now, the learner’s output for ‘adult form’ is ‘simplified form’, which is still an error. An additional winner-loser pair is added to the Support, as in (20).

(20)

	M ₁	M ₂	M ₃	F ₁	F ₂	F ₃	MP(M ₁)	MP
adult > ∅	L						W	W
adult > simplified	L			W				

The Support in (20) has three constraints that could be installed: F₁, MPARSE(M₁), and MPARSE. We install F₁ first due to the F > MPARSE bias. Next, MPARSE(M₁) is

chosen for installation, presumably because it is more specific than MPARSE. The resulting grammar, in (21), allows forms that violate M_1 to surface faithfully, but forms that do not violate M_1 are still avoided.

- (21) \mathcal{H}_2 : $M_2, M_3, \dots \gg F_1 \gg \text{MPARSE}(M_1) \gg M_1 \gg F_2, F_3, \dots \gg \text{MPARSE}$
 M_1 -violating forms faithful; everything else avoided.

The learner achieved an adult-like grammar as far as M_1 is concerned. The acquisition path goes through avoidance and then simplification on its way to the adult form, all the while allowing forms that do not violate M_1 to remain avoided. Regarding the formal definition of the markedness-specific versions of MPARSE, it should be noted that $\text{MPARSE}(M_1)$ is defined relative to the fully faithful candidate (McCarthy 2003, 2007). In this stage of phonotactic acquisition, the fully faithful candidate (FFC) is conveniently identical to the adult form (as in, e.g. Hayes 2004).

- (22) $\text{MPARSE}(M_1)$: Assign a violation mark to \odot if the FFC violates M_1

The definition of the markedness-specific MPARSE can be made more general, applying to a set of markedness constraints.

- (23) $\text{MPARSE}(M_m \dots M_n)$: Assign a violation mark to \odot if the FFC violates all of the Markedness constraints in $M_m \dots M_n$

To summarize, I showed that the general MPARSE is too blunt as a tool for modeling avoidance in acquisition, as it creates unintended and implausible effects on the learner's behavior. Relativizing MPARSE to the markedness constraints that cause avoidance creates a well-behaved and realistic acquisition path, and offers a first formalization for Kiparsky & Menn's insight.

An anonymous reviewer asks whether it would be possible to relativize faithfulness constraints instead of MPARSE. While it is technically possible, it would lead to a much more complicated theory: to prevent avoidance in response to M_1 , F_1 will need to be demoted below MPARSE, but only for those cases where M_1 is involved. So F_1 will have to be split into one version that is sensitive to M_1 (to be demoted below MPARSE), and one version that is sensitive to all candidates that do not violate M_1 (to be kept above MPARSE). The general F_1 is removed from the grammar. Worse still, all other faithfulness constraints will also have to be specified as not applying to forms that violate M_1 . These negative definitions of faithfulness constraints will have to be repeated for each markedness constraint that causes simplification rather than avoidance. Relativizing MPARSE, then, is the preferred solution.

5 Implementation

The issue of learning Optimality Theoretic grammars has received much attention over the years, with much of the theory highly formalized and rigorously proven to work, as in e.g. Tesar & Smolensky (1998, 2000). The current model adds a new component to the theory, and thus should be checked carefully to work as advertised. I implemented the learner computationally as a Perl script. The script and accompanying files are available at <http://becker.phonologist.org/shaxar/>. The learner is overall similar to the one in Becker & Tessier (2011), but with constraints defined as bona fide functions, as in Becker (2005).

First, I will show the learner in its basic form, and show how it generates the problem discussed in §4.1. Then, a mechanism for installing markedness-specific versions of MPARSE will be added, and a more realistic acquisition path will emerge.

The implementation I offer is categorical in two senses: it moves from one grammar to the next categorically (i.e. there is only one grammar operative at each point in time), and each grammar is categorical (all constraints are categorically ranked). This results in a simplified acquisition path relative to SR's actual productions. See Becker & Tessier (2011) for a noisier implementation of Error Selective Learning.

5.1 Installing the general MPARSE

The learner starts with a set of four universal constraints (CON): *SONCODA, *COMPLEXONSET, MAX, MPARSE. When given an input, EVAL creates up to five candidates: the fully faithful candidate and \odot are always created. If the input has an initial cluster or a final sonorant, simplified forms that repair either or both are created. Given a candidate set, EVAL simply chooses the winner given the current ranking.

Whenever the learner produces a non-adult form, or an error (i.e. the fully faithful candidate is not the winner), it is stored in the Cache (Tessier 2007, 2008, 2009; Becker & Tessier 2011). When errors accumulate in the Cache, they are selectively moved to the Support. Then, the ranking algorithm (à la Prince & Tesar 2004; Hayes 2004) creates a new grammar, persistently favoring $M > F > \text{MPARSE}$.

As its input, the learner runs a list of SR's targets through its current grammar, one by one, in random order. Randomizing the list assures that the learner attempts words with marked structures throughout the learning process.

When enough errors accumulate, the Support is updated and a new grammar is learned. The learner starts with the initial state (\mathcal{H}_0) and passes through three intermediate grammars on its way to the adult grammar (\mathcal{H}_4).

As seen in Figure 3, the learner starts by being completely silent in \mathcal{H}_0 , as everything is avoided. Then, final sonorants are simplified in \mathcal{H}_1 , and at the same time, initial clusters are simplified as well. Avoidance is switched off at \mathcal{H}_1 for both marked structures. When $*\text{SONCODA}$ is ranked below faithfulness in \mathcal{H}_2 , final sonorants are pronounced faithfully, and initial clusters are avoided again. Initial clusters are again simplified in \mathcal{H}_3 , and produced faithfully in \mathcal{H}_4 . The appearance of simplified initial clusters in \mathcal{H}_1 is due to the installation of MPARSE above faithfulness, and their disappearance in \mathcal{H}_2 is due to the renewed installation of MPARSE below faithfulness. The learner adopts both of these grammars in an attempt to master final sonorants; the effect on initial clusters is unintended.

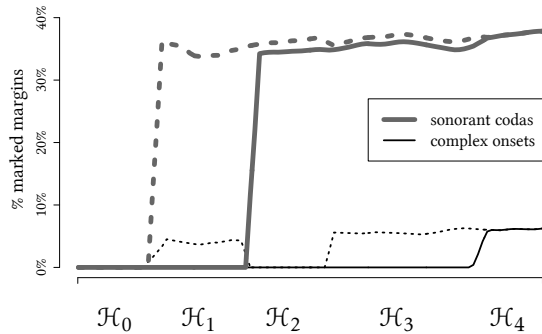


Figure 3: Model predictions: Productions and attempts of marked syllable margins. Solid lines represent faithful productions, dashed lines simplified productions.

It should be noted that the learner chooses what to learn from based on the accumulation of errors in the Cache, which is a buffer that determines which errors are sent to the Support. In Hebrew, final sonorants are much more common than initial clusters, which is why the learner masters them earlier. It would seem, however, that the frequency effect here is rather large: initial clusters are about five times rarer than sonorant codas, yet SR acquires them in fairly rapid succession. To mimic SR's pace, I allowed learning thresholds to differ by markedness constraint; in essence, making the model more sensitive to violations of $*\text{COMPLEXONSET}$ than frequency would require. While this improves the fit of the model to the data, further research will be needed to assess the appropriateness of this solution more generally.

5.2 Installing markedness-specific MPARSE

To fix the problem with the model in §5.1, a new model was built that installs markedness-specific versions of MPARSE , as defined in (23). The original MPARSE

stays at the bottom of the hierarchy.

Markedness constraints and MPARSE are functions from a linguistic form (a candidate) to a number of violations. Similarly, faithfulness constraints are functions from a pair of linguistic forms (the input and a candidate) to a number of violations. The new markedness-specific MPARSE proposed here is a little more involved, though perfectly well-defined: it is a function from a linguistic form (the candidate), a markedness constraint, and the fully faithful candidate, to a number of violations. The simulation creates these new constraints/functions on the fly, and they are added to CON .

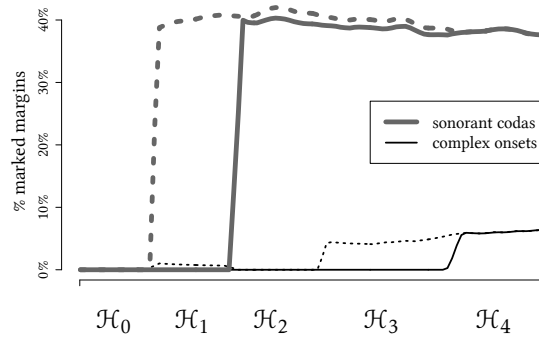


Figure 4: Model predictions: Productions and attempts of marked syllable margins

Running the simulation with this new mechanism in place makes the acquisition path more plausible, as show in Figure 4. In \mathcal{H}_1 , the general MPARSE is still at the bottom of the hierarchy, but $\text{MPARSE}(*\text{SONC})$ is installed above $*\text{SONCODA}$. Forms with initial clusters are still avoided, since the general MPARSE is below faithfulness. And yet, Figure 4 shows a few forms with initial clusters sneaking in during \mathcal{H}_1 . Why is that? The reason is that some words contain both an initial cluster and a final sonorant, and these words are not avoided when \mathcal{H}_1 is in place. Recall that simplifying an initial cluster or a final sonorant is done by deletion, so the same MAX-C is violated in both cases. When a form such as [pɸa.'χim] ‘flowers’, which has both marked structures, is run through \mathcal{H}_1 , the null parse is not the winner (24).

(24)

/pɸa.'χim/	*COMPLEX ONSET	$\text{MPARSE}(*\text{SONC})$	*SONCODA	MAX-C	MPARSE
a. pɸa.'χim	*!		*		
b. pa.'χim			*!	*	
c. ☞ pa.'χi				**	
d. ☉		*!			*

The emerging result is that two marked configurations that involve the same faithfulness constraint are predicted to interact, but *only in the words that have both configurations*. I find the prediction to be intriguing, yet SR's data is not sufficiently rich to prove or disprove it. Hopefully, more generous avoidance patterns will be identified and examined in this light in the future.

6 Conclusions

I analyzed SR's acquisition of marked syllable margins, showing that he first *avoids* words that have them, then he *deletes* the offending consonants, then produces them faithfully.

I offered an analysis of SR's acquisition path in terms of Error Selective Learning, modeling the avoidance of marked structure as selection of the null parse. The model includes a persistent $M > F > M_{\text{PARSE}}$ bias and relativization of M_{PARSE} to markedness constraints.

A result that emerges from the null parse approach to target selection is the dual utility of target selection: Firstly, selection removes non-adult-like productions, leaving the forms that the child does produce more adult-like than they would otherwise be. This makes the child's speech more likely to be interpreted correctly by hearers. Secondly, selecting the null parse is informative: it identified the markedness constraints that are involved, and it causes the learner to transition into simplification and then into adult-like speech.

The analysis is implemented computationally, which serves as a check on its accuracy and its ability to generate realistic acquisition paths. The selection of constraints to install is a rather complex procedure, as it goes through markedness, faithfulness, and M_{PARSE} , treating each one differently. It would seem that the mechanism can be improved and streamlined, perhaps with an optimization technique on the selection process; this is left for future work.

Questions remain about the general applicability of the approach taken in this paper. I diagnosed target avoidance in SR's corpus by looking at his intended productions only. The assumption that two different kinds of marked structures are repaired in the same way relies additionally on the transcriptions of his productions, which requires a greater reliance on the accuracy of the corpus. It is hoped that future research will shed more light on the repairs that children make, and improve our confidence in analyses that rely on repair identification.

Going beyond SR's case, questions remain about the kinds of structures that children avoid, and differences between children in their tendencies to avoid certain structures and not others. Assuming that such differences do exist, the

learning mechanism would need to accommodate these differences, hopefully by nothing more than minor adjustments to the model's parameters.

References

- Adam, Galit & Outi Bat-El (2007a). Segmental effects on syllable selection: Evidence from Hebrew. Handout from Generative Approaches to Language Acquisition, Barcelona.
- Adam, Galit & Outi Bat-El (2007b). The trochaic bias is universal: Evidence from Hebrew. Handout from Generative Approaches to Language Acquisition, Barcelona.
- Adam, Galit & Outi Bat-El (2009). When do universal preferences emerge in language development? The acquisition of Hebrew stress. *Brill's Annual of Afroasiatic Languages and Linguistics* 1. 255–282.
- Adam, Galit & Outi Bat-El (2010). The trochaic bias in the acquisition of Hebrew. Ms. Tel Aviv University.
- Albright, Adam (2010). Lexical and morphological conditioning of paradigm gaps. In *Modeling Ungrammaticality in Optimality Theory*, London: Equinox.
- Albright, Adam (2011). Gradient complexity thresholds in phonology. Talk given at Stanford University.
- Baayen, R. Harald (2008). *Analyzing Linguistic Data: A practical introduction to statistics*. Cambridge University Press.
- Bat-El, Outi (2010). A gap in the feminine paradigm of Hebrew: A consequence of identity avoidance in the suffix domain. In Curt Rice (ed.) *When Nothing Wins: Modelling Ungrammaticality in OT*, London: Equinox.
- Becker, Michael (2005). CCamelOT - an implementation of OT-CC's GEN and EVAL in Perl. Software, UMass Amherst, <http://people.umass.edu/mbe/CCamelOT/>.
- Becker, Michael & Anne-Michelle Tessier (2011). Trajectories of faithfulness in child-specific phonology. *Phonology* 28.
- 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.
- Goad, Heather & Yvan Rose (2004). Acquisition of left-edge clusters in West Germanic. In René Kager, Joe Pater & Wim Zonneveld (eds.) *Constraints in Phonological Acquisition*, Cambridge University Press. 109–157.
- Grijzenhout, Janet & Sandra Joppen-Hellwig (2002). The lack of onsets in German child phonology. In I. Lasser (ed.) *The Process of Language Acquisition*,

- Frankfurt/Berlin: Peter Lang Verlag. 319–339.
- 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.
- Jaynes, E. T. (2003). *Probability Theory: The Logic Of Science*. Cambridge University Press.
- Kager, René, Joe Pater & Wim Zonneveld (2004). *Constraints in Phonological Acquisition*. Cambridge University Press.
- Kiparsky, Paul & Lise Menn (1977). On the acquisition of phonology. In *Language Learning and Thought*, New York: Academic Press. 47–78.
- McCarthy, John J. (2003). Comparative markedness. *Theoretical Linguistics* 29. 1–51.
- McCarthy, John J. (2007). *Hidden Generalizations: Phonological Opacity in Optimality Theory*. London: Equinox.
- McCarthy, John J. & Matthew Wolf (2010). Less than zero: Correspondence and the null output. In Curt Rice (ed.) *Modeling Ungrammaticality in Optimality Theory*, London: Equinox.
- Menn, Lise (2004). Saving the baby: making sure that old data survive new theories. In René Kager, Joe Pater & Wim Zonneveld (eds.) *Constraints in Phonological Acquisition*, Cambridge University Press. 54–72.
- Mudelsee, Manfred (2010). *Climate Time Series Analysis*. Springer.
- Operstein, Natalie (to appear). Mapping the linguistic sub-areas of Eurasia: Word-initial /r-/. In Vitali V. Shevoroshkin & Harald Sverdrup (eds.) *Festschrift in Honor of Aharon Dolgopolsky*.
- 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 (2011). R: A language and environment for statistical computing. Vienna, Austria. (<http://www.R-project.org>).
- Rasmussen, Jens Elmegård (1999). New evidence for infixal -o- in Indo-European: Ablaut re-/or-. In *Selected Papers on Indo-European Linguistics*, Copenhagen: Museum Tusulanum Press, University of Copenhagen, vol. 1. 327–335.
- Rice, Curt (2006). Nothing is a phonological fact: Gaps and repairs at the phonology-morphology interface. Talk given at NELS 36.
- Schwartz, Richard G. (1988). Phonological factors in early lexical acquisition.

- In Michael D. Smith & John L. Locke (eds.) *The Emergent Lexicon: The Child's Development of a Linguistic Vocabulary*, Academic Press, New York, chap. 7. 185–222.
- Schwartz, Richard G. & L. Leonard (1982). Do children pick and choose? an examination of phonological selection and avoidance in early lexical acquisition. *Journal of Child Language* **9**. 319–336.
- Schwartz, Richard G., L. Leonard, D. Loeb & L. Swanson (1987). Attempted sounds are sometimes not: An expanded view of phonological selection and avoidance. *Journal of Child Language* **14**. 411–418.
- Smith, Brian (2009). Cumulative constraint interaction and the null parse. Talk given at HUMDRUM '09 and MUMM '09.
- Stoel-Gammon, Carol & Judith A. Cooper (1984). Patterns of early lexical and phonological development. *Journal of Child Language* **11**. 247–271.
- Tesar, Bruce & Paul Smolensky (1998). Learnability in Optimality Theory. *Linguistic Inquiry* **29**. 229–268.
- Tesar, Bruce & Paul Smolensky (2000). *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.
- Tessier, Anne-Michelle (2006). Testing for OO-Faithfulness in artificial phonological acquisition. In David Bamman, Tatiana Magnitskaia & Colleen Zaller (eds.) *Proceedings of BUCLD30*, Cascadia Press. 619–630.
- Tessier, Anne-Michelle (2007). *Biases and stages in phonological acquisition*. Ph.D. dissertation, University of Massachusetts Amherst, Amherst, MA.
- Tessier, Anne-Michelle (2008). Children's exceptional words: Lexical challenges from the time-course of L1 phonological acquisition. Talk given at the University of Western Ontario.
- Tessier, Anne-Michelle (2009). Frequency of violation and constraint-based phonological learning. *Lingua* **119**. 6–38.