

# U-Shaped Learning in Phonological Development

Joseph Paul Stemberger  
University of Minnesota

Barbara Handford Bernhardt

Carolyn E. Johnson  
University of British Columbia

*This is a lightly edited version of a poster that was presented at a conference, the citation for which is:*

Stemberger, J.P., & Bernhardt, B.H., & Johnson, C.E. (1999). "Regressions" ("u"-shaped learning) in the acquisition of prosodic structure. Poster presented at the 6<sup>th</sup> International Child Language Congress, July 1999.

## 0. Introduction

This paper addresses learning of constraint rankings by children. We argue that constraint demotion cannot account for all developmental patterns observed in child phonology. Learning algorithms that allow for constraint promotion fare better, but some still cannot account for all developmental changes.

Child language contains many inaccuracies relative to the target adult language. Over the course of development, these inaccuracies disappear. Most developmental changes bring the child's speech closer to the target language. Theories of learning focus on how such improvements are accomplished.

However, a minority of developmental changes are **U-SHAPED**:

a developmental change in which there is a decrease in accuracy relative to the adult target.

Historically, much of the debate over different theories of learning has focused on how each theory deals with these relatively rare developmental occurrences (e.g. Macken, 1992; Pinker & Prince, 1988; Stemberger & Bernhardt, 1998). Regressions are also the best test of the adequacy of learning models proposed for OT. We argue that constraint promotion must be a possible change during

learning, contra Tesar and Smolensky (1998), and Prince and Tesar (1999). Some learning algorithms that allow for constraint promotion, such as that of Boersma and Hayes (1999), may nonetheless have difficulty accounting for some regressions, because of restrictions on which constraints may be promoted. These models fail to account for the particular set of regressions examined here, and hence are not adequate to describe learning by human beings.

## 1. A Typology of Phonological Regressions

U-shaped learning appears to be especially prevalent in the development of phonology (the pronunciation of words), where it is known as *regression*. We have found it useful to describe phonological regressions using two dimensions that describe how the system has changed:

### **TRADE-OFF vs. PURE:**

whether along with the regression there is some improvement *in the same output*

### **CORRELATED vs. UNCORRELATED:**

whether the regression correlates with a change made to the system that leads to improvements *in other outputs* than the one showing the regression

*ALL* theories of learning predict regressions only when there is a motivation:

**trade-off:** Due to limitations in the range of outputs, altering the system to achieve accuracy on one dimension of a form may be at the expense of accuracy on another dimension.

**correlated:** Changes that increase the range of possible outputs or that improve the efficiency of processing may make errors possible that before were impossible, so the system was not set up to prevent such errors from occurring.

These two factors predict four basic types of regression:

correlated trade-off  
uncorrelated trade-off  
correlated pure  
uncorrelated pure

Theories of learning do not predict **uncorrelated pure** regressions, where forms get worse for no apparent reason.

### ***1.1 Restrictions on the types of re-ranking that are possible***

Different theories of learning make different predictions about the sorts of trade-off regressions that are possible. Tesar and Smolensky (1998) and Prince and Tesar (1999) assume that constraints may only be demoted, never promoted. For example, given an incorrect ranking **NegativeOutputConstraint** » **FaithfulnessConstraint**, the negative output constraint can be demoted (re-ranked lower), but the faithfulness constraint cannot be promoted (re-ranked higher). Whether a constraint is promoted or demoted makes a practical difference only when a third constraint is originally ranked between these two constraints:

*original ranking:*

**NegativeOutputConstraint** » **ThirdConstraint** » **FaithfulnessConstraint**

*demotion of highest constraint:*

**ThirdConstraint** » **NegativeOutputConstraint** » **FaithfulnessConstraint**

*promotion of lowest constraint*

: **NegativeOutputConstraint** » **FaithfulnessConstraint** » **ThirdConstraint**

Either re-ranking can in theory lead to a trade-off regression: a new type of error occurs along a phonological dimension that had previously been output correctly. Bernhardt and Stemberger (1998) and Stemberger and Bernhardt (1998) have argued that some regressions in child phonological development clearly involve promotion.

Boersma and Hayes (1999) develop a different theory of learning, in which both demotion and promotion occur. The two incorrectly ranked constraints are identified. The higher-ranked constraint is demoted slightly in the full hierarchy of constraints, while the lower-ranked constraint is promoted slightly. Over many learning trials, the two constraints are ranked closer and closer together, until eventually the ranking is reversed. They also assume that the exact ranking of constraints is subject to slight variability on each output trial, so that variability of output can result. This leads to the prediction that incorrect outputs do not disappear all at once, but co-vary for some period with the new form (the correct form, or a new error). Stemberger and Bernhardt (1998) note that such variability is the hallmark of changes in phonological development, so this aspect of Boersma and Hayes' model is very desirable.

However, Boersma and Hayes (1999), along with Tesar and Smolensky (1998) and Prince and Tesar (1999), posit a strong restriction on the possible re-rankings that may occur. The first step in learning identifies the constraints that are ranked incorrectly. Only incorrectly ranked constraints may be re-ranked. This is not as obviously necessary as it may seem. There are instances in which the re-ranking of an already-correctly-ranked constraint could eliminate a particular type of error (at the expense of introducing another type of error). For example, if /l/ undergoes gliding to [j], this error could be eliminated by promoting constraints against glides: if [j] is not a possible output, then /l/ cannot undergo gliding. However, glides are possible in adult English, so promoting the constraints against

glides is a change in the wrong direction; the low ranking of anti-glide constraints in the child's system is not where the real problem lies. However, in a learning system that is free to re-rank any constraint that would lead to an improvement along the dimension that was in error, at the cost of a trade-off regression, promoting the anti-glide constraints would be a possible way to eliminate that type of error. We argue that one regression discussed here is of this sort.

## 2.0 THIS STUDY

We examine a complex set of regressions in the speech of one child learning English. We argue that changes made to some vowel-final words over-generalized in a complex way to cause regressions on all vowel-final words. The regressions appears to involve the "incorrect" promotion of a constraint that was correctly low-ranked, as part of a trade-off to improve faithfulness to [Labial].

### Subject

Subject was one female English-learning child, who was part of a longitudinal study from 0;11 through about 4;0. Subject was Morgan, the second daughter of the first author.

### Method

A classic diary-study methodology was used, as in Smith (1973). The child's speech was followed carefully. Notebooks were kept on hand to record in writing interesting aspects of the child's utterances. An attempt was made to record every new word in narrow transcription, plus every variant pronunciation and new pronunciation of old words. Words were often transcribed even though no change had occurred, just to document the pronunciations. When two-word and larger sentences appeared, these were also transcribed.

## 2.1 Beginning State of The Child's System

1. Final consonants were present from 0;11, and were *never* subject to deletion.
2. When a word ended in a vowel, the consonant [ʔ] was optionally epenthesisized, along with the shortening of the long-vowels /uː/, /iː/, /oː/:

*No.*    /noː/    [nʌ:] ~ [nʌʔ]

\*        in utterance-final stressed syllables        (only one record after 1;4.1)

\*        in word-final unstressed syllables        (optional throughout the period here)

*Hungry.*        /hʌŋgɹi/        [hʌŋki] ~ [hʌŋkiʔ]

\* in word-final stressed syllables in the first word of a two-word utterance

*No bonking!* /no: bɑŋkiŋ/ [nʌ? bɑŋkiŋ]

3. There were no rounded vowels.  
The [Labial] feature of rounded vowels was simply deleted, yielding the equivalent unrounded vowel.

<i>No.</i>	/no:/	[nʌ:]
<i>Shoe.</i>	/ʃu:/	[ʧʊ:]
<i>Cow.</i>	/kaʊ/	[tʰaʊ]

## 2.2 Regression 1: *Vowel Rounding* 1;6.22

	<u>Point 1</u>	<u>Point 2</u>
<i>No!</i>	nʌ:	nʌ: ~ nʌ:m
<i>Me!</i>	mi:	mi:
<i>Mouse.</i>	maʊ?	maʊ?

*change:* Word-final [m] was *optionally* inserted after vowels that should have been rounded, **but only before a pause.**

*motivation:* The vowel's [Labial] feature was expressed.

*exceptions:* *Yellow.* [la:lʌ:] ~ [la:lʌ?] (never \*[la:lʌm])

Many low-frequency words were never observed with final [m], but it may be that [m]-final variants were possible, but did not occur during the observation period.

<i>chain-shift:</i>	<i>Zoom!</i>	[jʊ:n]
	<i>Gnome.</i>	[nʌ:n]

It is clear that the child had been perceiving the difference between words that ended in rounded vowels vs. words that ended in /m/, from the onset of meaningful speech. Word-final /m/ was **always** realized as [ʔ] from the very first time it was attempted, whereas vowel-final words varied between being vowel-final and ending in [ʔ]. By 1;4, final /m/ was **always** realized as [n], while vowel-final words were generally vowel-final, and only occasionally ended in [ʔ], never in a nasal. This regression was thus not motivated by an earlier failure to perceive the difference between lip-rounding and a final

[m]. (Macken, 1992, has suggested that regressions may sometimes be rooted in earlier misperceptions and/or inaccurate underlying forms.)

<i>over-generalization:</i>	<i>Sockie.</i>	[ lakim ]
	<i>Doggy.</i>	[ dagim ]
	<i>Me!</i>	[ mi:m ]

*type:* **TRADE-OFF**

<i>improved:</i>	vowel's [Labial]
<i>worsened:</i>	vowel-final-ness

**UNCORRELATED**

**2.3 Regression 2: Word-Final [h] 1;7.0**

	<u>Point 2</u>	<u>Point 3</u>
<i>No!</i>	nʌ: ~ nʌ:m	nʌ: ~ nʌ:m ~ nʌ:h
<i>Me!</i>	mi:	mi: ~ mi:h
<i>Mouse.</i>	maʊ?	maʊh

*change:* Word-final /s/ was now realized as [h].

*motivation:* The [+continuant] feature of /s/ was expressed.

<i>exceptions:</i>	<i>Geese.</i>	/ gi:s /	[ di:ʔ ]
	<i>Nurse.</i>	/ nɜ:s /	[ nʌ:ʔ ]

<i>over-generalization:</i>	<i>Me!</i>	[ mi:h ]
	<i>Ball.</i>	[ baʊh ]
	<i>No!</i>	[ nʌ:h ]

**Over-generalization was UTTERANCE-FINAL ONLY.**

*type:* **PURE**

<i>worsened:</i>	vowel-final-ness
------------------	------------------

**CORRELATED**

with improvement to final /s/

notes: **UNLIKE** earlier pronunciations with [ʔ], the vowel never shortened.

## 2.4 Regression 3: *Word-Final [ə]* 1;7.5

### Point 4

<i>No!</i>	nʌ: ~ nʌ:m ~ nʌ:h
<i>Me!</i>	mi: ~ mi:h ~ mi:ə
<i>Mouse.</i>	maʊh

*change:* A [ə] was optionally epenthesized after word-final /i/ and /aɪ/.

*motivation:* NONE APPARENT.

*type:* **PURE**

*worsened:* [ə] added to end of word

??? **CORRELATED** ???

with other vowel-final regressions

*Note 1:* Almost always **UTTERANCE-FINAL**.  
(There was a single record in the first word of a two-word utterance.)

*Note 2:* Why just after /i/?

- \* After [ʌ], [a]: dissimilation of vowel height?  
(Two non-high vowels disallowed)
- \* After [ɪ], [aɪ]: perhaps [Labial] always spread to a following voiced segment, forcing it to be [m]

## 3.0 THE CHALLENGE

How can these regressions be explained? Theories of acquisition have **no obvious explanation for uncorrelated pure regressions**. If a theory can treat Regressions 2 and 3 as correlated pure regressions, we regard it as a superior theory.

### 3.1 *Process/Rule Analysis*

Regression 2:

With the change of /s/ from [ʔ] to [h], the processes would be:

Point 2:  $s \rightarrow ? / \text{ \_\_\_ } C \text{ \_\_\_ } ]_{\text{word}}$

Point 3:  $s \rightarrow h / \text{ \_\_\_ } C \text{ \_\_\_ } ]_{\text{word}}$

However, since there is no /s/ at the end of vowel-final words, this process could not generalize. New and unrelated processes would be needed:

$\emptyset \rightarrow h / V \text{ \_\_\_ } ]_{\text{word}}$

$\emptyset \rightarrow \text{ə} / V \text{ \_\_\_ } ]_{\text{word}}$   
 +hi  
 -bk

There is no way to relate these processes to any other process. Regressions 2 and 3 are UNCORRELATED PURE regressions. This is an undesirable result.

### 3.2 Connectionist Models

Vowel-final forms were statistically marginal in the child's speech. **Type frequency** in adult words used in the child's speech at 1;6.21:

	Final Stressed Syllable	Final Unstressed Syllable
Word ends in Consonant	128	38
word ends in Vowel	37	50
% C-final	77.6%	43.2%

Regression 1 led to a DECREASE in the frequency of vowel-final forms. Words that before were always vowel-final were often produced as consonant-final ( $[n\text{ʌ}:\text{ɪ}] \rightarrow [n\text{ʌ}:\text{m}]$ ). There was an increase in the transitional probability of  $V \rightarrow C$ .

The system was already inclined to produce VC in unstressed syllables and in the first word in a sentence. This makes the system prone to adding nonsyllabic elements at the end of the word. However, the lexical item does not supply any features, so the features within the inserted segment must come from somewhere else. A common solution for supplying the features (in phonological development, and also in adult languages) is the following (Bernhardt & Stemberger, 1998):

**Derive some or all features via assimilation**

**Possibility 1** (*Regression 2*):

Spread the feature [+continuant] from the vowel.

At that point in development, the only possible word-final [+continuant] phone was [h].

**Possibility 2** (*Regression 3*):

Spread the features [+continuant] and [+voiced] from the vowel.

At that point in development, the only possible word-final phone that was  
[+continuant,+voiced] was a vowel.

The default vowel was [ə] (mid central).

Regression 3 may be the result of trying to correct Regression 2. If the insertion of the non-lexical [-voiced] voicing feature can be prevented, then [h] will not result. However, this merely changes the nature of the error, since a segment can still be inserted if [+voiced] can be spread from the preceding vowel. (That re-ranking itself implies the *promotion* of the negative constraint **Not(+voiced)** (or **DEP(+voiced)**, if that constraint is preferred), and so is problematic for learning theories that allow only constraint demotion.)

In a Connectionist approach, Regression 2 may result from a decrease in the frequency of vowel-final forms (due to Regression 1), in combination with the recent change that made [h] a possible word-final phone. Regression 3 can be similarly motivated, and may additionally have arisen as part of the child's attempt to correct Regression 2.

Regressions 2 and 3 can be viewed as CORRELATED PURE regressions, which can be motivated as an expected type of developmental change.

### 3.3 *Optimality Theory*

In Optimality Theory (e.g. Prince & Smolensky, 1993; Bernhardt & Stemberger, 1998), all outputs are determined via the relative ranking (or importance) of constraints that can have motivated exceptions (and that thus constitute tendencies rather than absolutes). Change occurs when constraints are re-ranked: when the relative importance of different constraints changes.

From the beginning of this child's system, rounded vowels were impossible, probably as a result of constraints on complexity. The back-tongue articulation that is part of the defining characteristic of vowels could not be combined with a lip articulation:

**NotComplex(Place) » Survived(Labial)**

The expected change is for these two constraints to simply reverse ranking:

**Survived(Labial) » NotComplex(Place)**

However, there is another way to get the lip articulation to be present in the pronunciation: to create a new segment and shift [Labial] into it:

*original ranking:*

**NotComplex(Place) » Not(C-Root) » Survived(Labial)**

*re-ranking:*

**NotComplex(Place) » Survived(Labial) » Not(C-Root)**

This ranking is *NOT* the correct one for adult English, and so yields a trade-off regression. However, more must be involved than this, because this cannot generalize to give Regressions 2 & 3.

A second alternative account posits that the rounded vowel is split into segments: an unrounded vowel, plus [m]. This would be a violation of **NoMultipleCorrespondence**. If this were initially ranked between **NotComplex(Place)** and **Survived(Labial)**, the promotion of **Survived(Labial)** would lead to the splitting of e.g. /o/ into [ʏ] and [m].

*original ranking:*

**NotComplex(Place) » NoMultipleCorrespondence. » Survived(Labial)**

*re-ranking:*

**NotComplex(Place) » Survived(Labial) » NoMultipleCorrespondence.**

However, there would be no way to generalize this to words like /mi:z/, in which all features were already accurately produced. Regressions 2 and 3 would have to be treated as uncorrelated pure regressions.

There is a third alternative that works. Bernhardt & Stemberger (1998) introduce the following constraint:

**WordFinalMassiveness:** Words end with extra *non-prominent* elements.

They suggest that this is a high-level phonological instantiation of Word-Final (and Phrase-Final) Lengthening:

The duration of syllables increases at the end of a word and at the end of a phrase, especially before a pause.

In this particular instance, this constraint forces the presence of a word-final consonant, and so is equivalent (for this particular instance) to the constraint **Final-C** that requires words to end in a consonant (McCarthy & Prince, 1994).

We suggest that, in order to create a word-final segment to take the [Labial] feature of the vowel, this constraint was re-ranked higher than it should be in English:

**WordFinalMassiveness » Not(Root)**

This would have the desired result of creating a segment to take the [Labial] of the vowel, but would have the negative side-effect of tending to create non-syllabic elements at the end of *any* vowel-final word.

Thus, Regressions 2 & 3 were generalizations of the changes that led to Regression 1. The rest of the explanation is the same as that given above for Connectionist models.

The promotion of this constraint is itself problematic, however. The correct ranking for adult English is a low ranking. The constraint preventing the epenthesis of an additional final segment (**Not(Root)**) *must* be ranked lower than **WordFinalMassiveness**, or else all English words would be required to end in consonants. Neither of these constraints were incorrectly ranked. The learning algorithms of Tesar and Smolensky (1998), Prince and Tesar (1999), and Boersma and Hayes (1999) all predict that neither of these constraints should be affected in re-ranking constraints to allow vowels to be rounded. Tesar and Smolensky, and Prince and Tesar, cannot account for Regression 1 at all. Boersma and Hayes, in contrast, can only account for Regression 1 given the following ranking at Point 1:

**NotComplex(Place) » Not(C-Root) » Survived(Labial)**

As **Survived(Labial)** is re-ranked higher during learning, it can come to be ranked higher than **Not(Root)**, so that a final consonant can be epenthesized to take on the vowel's [Labial] feature, thereby allowing the feature to survive in the output. However, as noted above, there is no way for this analysis to also derive Regressions 2 and 3, which would have to be treated as uncorrelated pure regressions, a type of regression for which Boersma and Hayes (and everyone else) have no explanation.

The re-ranking that actually occurred involved the promotion of a constraint that was correctly low-ranked. This re-ranking solved the immediate problem: the vowel's [Labial] feature was now present in the output. However, it was present in an unusual fashion that represented a decrease in faithfulness to the adult target along another dimension. It was a classic trade-off regression. Since this was the wrong re-ranking, however, it was ultimately re-ranked again.

#### 4.0 Point 5: *An increase in accuracy*

1;8.10

<i>No!</i>	nʌ:
<i>Me!</i>	mi:

*Mouse.*

mauh

For 7 weeks, the vowel feature [Labial] had optionally been expressed, but in the wrong way. By re-ranking **WordFinalMassiveness** lower (as in adult English), this ended, leaving vowel rounding the way it had begun: simply being deleted.

Eight months later, at 2;6, constraints were re-ranked into an approximation of adult English, and rounded vowels occurred in an adult-like fashion.

## 5.0 CONCLUSIONS

Regressions (u-shaped learning) generally have a clear motivation in phonological development:

- \* There is often a trade-off involved:  
one aspect of the pronunciation of a word improves while another worsens.
- \* There is often a correlation with a change elsewhere in the system. The change makes the system prone to types of error that before were not possible.

Uncorrelated pure regressions are rare, if they even occur at all. No theory of learning that we are familiar with has an explanation for them.

We have examined a set of regressions from the speech of one English-learning child. We have argued that these regressions can be viewed as correlated pure regressions in Connectionist Models and in Optimality Theory, but not in Process Theory. We take this as evidence for the superiority of Connectionist Models and of Optimality Theory.

However, not all OT learning models can account for all details of the regressions. Models that rely entirely on constraint demotion (Tesar & Smolensky, 1998; Prince & Tesar, 1999) are inadequate. Boersma and Hayes (1999) fare better because constraint promotion is always a part of change in their model. However, they place restrictions on which constraints may be re-ranked that are not true of learning in young children. Prince and Tesar (1999: 4) state that the “proposals we will entertain are not intended to provide a direct account for child language data, although we expect that they ought to bear on the problem in various ways.” We have argued that their proposals do not provide a full account of phonological development. Current computationally-explicit theories of learning in OT may be adequate as theories of machine learning, but clearly fail if viewed as a theory of the learning of human language by humans.

## References

Bernhardt, B.H., & Stemberger, J.P. (1998). *Handbook of phonological development: From the perspective of constraint-based nonlinear phonology*. Academic Press: San Diego, CA.

- Boersma, P., & Hayes, B. (1999). Empirical tests of the Gradual Learning Algorithm. ROA-348.
- Macken, M.A. (1992). Where's phonology? In C.A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, and implications* (pp. 367-397). Hillsdale, NJ: Lawrence Erlbaum Associates.
- McCarthy, J. J., & Prince, A. S. (1994). The emergence of the unmarked: Optimality in prosodic morphology. *Proceedings of the North Eastern Linguistics Society*, *24* (pp. 333-379). GSLA, University of Massachusetts, Amherst.
- Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a Parallel Distributed Processing model of language acquisition. *Cognition*, *28*, 73-194.
- Prince, A. S., & Smolensky, P. (1993). Optimality theory: Constraint interaction in generative grammar. (Rutgers University Cognitive Sciences Center Technical Report-2). Piscataway, New Jersey.
- Prince, A., & Tesar, B. (1999). Learning phonotactic distributions. ROA-353.
- Smith, N. (1973). *The acquisition of phonology*. Cambridge: Cambridge University Press.
- Stemberger, J.P., & Bernhardt, B.M. (1998). The emergence of faithfulness. In B. MacWhinney (Ed.), *The emergence of language* (pp.417-446).Mahwah, NJ: Lawrence Erlbaum Associates.
- Tesar, B., & Smolensky, P. (1998). The learnability of Optimality Theory. *Linguistic Inquiry*, *29*, 229-268.