U-shaped learning in language acquisition, and restrictions on error correction

Joseph Paul Stemberger* & Barbara Handford Bernhardt
University of British Columbia
* *University of Minnesota at the time of the conference*

Poster presented at the 2001 Biennial Meeting of the Society for Research in Child development, April 22,2001.
This is exactly as posted.

Children begin language acquisition with systems that produce many errors.

Error correction leads to a more adult-like system.

Each model of learning imposes particular characteristics on how change can occur.

- Some models allow any sort of change to occur.
- Other models strongly restrict the sorts of change that are possible.

This paper tests whether the restrictions imposed by one model (Tesar & Smolensky, 2000) are too severe. It does so by addressing instances in which change leads to a system that is in some respects a step away from the adult system. *Such U-shaped learning is often regarded as the strongest challenge for models of learning.*

Two normal developmental paths:

1. S-shaped: it just keeps getting better
   - Point 1: 0% correct
   - Point 2: variability between correct and incorrect
   - Point 3: (almost) 100% correct

2. U-shaped learning: it gets worse before it gets better
   - Point 1: 100% correct
   - Point 2: 50% correct
   - Point 3: (almost) 100% correct

Why is U-shaped learning interesting?

Many learning theories assume ERROR CORRECTION.

So, how does learning INDUCE error?

RESTRICTIONS ON CHANGE TO THE SYSTEM

Connectionist:

change is motivated by error-correction

change is also motivated by frequency of output independent of error
- an increasing frequency effect
- plus possibly some anti-frequency effects *{But that's another story.}*

otherwise, no restrictions on the nature of possible changes

predicts period of variability when changing the system

Connection weights are altered to eliminate error X from output.

This makes X a less likely output, and can lead to errors when X *should* be output.

Optimality Theory:

a connectionist-inspired symbolic model that eliminates rules and makes use only of **constraints**

Constraints are ranked.

High-ranked constraints are more constraining than low-ranked constraints.

Low-ranked constraints are violated rather than high-ranked constraints.

Differences in the ranking of two constraints leads to different output, in two different adult languages, or in two different child systems.

Bernhardt & Stemberger (1998):

Any change that eliminates the targeted error is possible, *as long as* the change is a **small** re-ranking; constraints can be re-ranked **higher** *or* **lower**.

Ranking is variable, and is correlated with activation levels.

This *predicts* the (observed) period of variability when changes are made.

Tesar & Smolensky (2000):

All constraints are strictly ranked.
A learner can only **DEMOTE** constraints, and only violated constraints can be demoted.

In an error, a constraint is violated by the correct pronunciation.
- A learner can tell that the constraint is ranked too high, if it rules out the correct output
- **DEMOTE** in **LARGE** steps (*Note: This leads to change that is much faster than observed in human language acquisition. It is necessary to adopt a variant in which all changes are small.*)

For low-ranked constraints that are violated by the target pronunciation, a learner can't tell which is the one that is poorly ranked.
It is therefore unclear which constraint to promote.
If the wrong constraint is promoted, it may not entirely solve the problem, and may cause errors. Tesar & Smolensky therefore reject constraint promotion as a mechanism of change.
This does not *predict* period of variability (but it can be stipulated).
(Symbolic models can ***always mimic*** *what connectionist models give for free.)*

Boersma & Hayes (1999):

Constraint ranking is variable; the rank of a given constraint varies higher or lower around a mean ranking value.

When there's an error, the learner identifies which two constraints are incorrectly ranked with respect to each other, and
- **decreases** the ranking of the higher-ranked constraint, *and*
- **increases** the ranking of the lower-ranked constraint

Only violated constraints can be actively re-ranked.
But actively re-ranking a constraint changes its ranking relative to other correctly-ranked constraints.

This *predicts* the period of variability when making changes to system.

We now examine an instance of U-shaped learning in first language acquisition that appears to involve promotion of constraints, contra Tesar & Smolensky.

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.

Methodology

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 every variant morphological form) 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.

Instances of U-shaped learning were identified. In on-going research, we are attempting to identify causal factors, in order to test alternative learning theories. For this paper, we identified an interesting interaction of morphology and phonology.

Beginning State of The Child's System (3;2)

Plurals and past tense forms:

A. **CORRECT**; words with target syllabic allomorphs

*young children often show **affix-checking** on the forms in (a) and (b), an error in which the base form is used, e.g. lift, horse. This child had **never** produced affix-checking errors before 3;2.*

a. past tense in bases ending in /t, d/: lifted [lɪftəd]
tasted [tʰeɪstəd]

b. plurals in bases ending in /s, z/: horses [hɔəsəz]
foxes [fɑksəz]

c. plurals in bases ending in fricatives: giraffes [ʤəʊæfs]
dishes [dɪʃəz]

B. **ERRORS**

a. variably, syllabic allomorph of past after /z/:

(*adult* [juːzd])	<u>used</u>	[juːzəd]	~	[juːd]		
(adult [ʧoːz]*)*	<u>choosed</u>	[ʧuːzəd]	~	[ʧuːd]	~	[ʧoːz]

b. syllabic allomorph of plural after /sC/: <u>ghosts</u> [goːtəz]
(*adult* [goːsts])

c. double-marking errors in past & plural: <u>trollses</u> <u>endses</u>
(*variable and infrequent*) <u>trippeded</u> <u>fixeded</u>

Regression I (3;3-3;4)

Errors corrected: All errors above disappeared.

Correctness maintained: In (Ac) above, the syllabic allomorph continued to be used after palatoalveolars.
The nonsyllabic allomorph continued to be used after other fricatives.

Errors induced: Affix-checking appeared *for the first time.*
It was not obligatory, but was observed on ca. 50% of all produced forms.
It appeared for both plurals and past tense forms.

Regression II (3;5)

Errors corrected: Affix-checking errors disappear, on both plurals and past tense forms.

Correctness maintained: The nonsyllabic allomorph continued to be used after /z/:
[juːzd]

Errors induced: Double-marking errors return, on both plurals and past tense forms. *Still variable but infrequent.*

Syllabic allomorph after fricative-stop clusters returns: ghosts [goːstəz]
Variable, but frequent. lifts [lɪftəz]

Occasional syllabic allomorphs after other fricatives: giraffes [ʤəʊæfəz]

Discussion

1. **Affix-checking** has commonly been observed. Children often fail to mark affixes after bases that end in similar segments (e.g., Berko, 1958; MacWhinney, 1978; Bybee & Slobin, 1981).

2. **Double-marking errors** have also been reported, and have been shown to have phonological conditioning for at least some children (Bernhardt & Stemberger 1998). (Although Pinker & Prince, 1988, suggest that there should be no phonological conditioning on double-marking errors, such conditioning is predicted by modern linguistic theories.) For this child, we see two possibilities:

(a) All observed errors involved monosyllabic nouns and verbs. The errors may have been motivated by a tendency for words to have two syllables.

(b) All such errors resulted in legal intervocalic clusters, e.g. /st/ and /nd/, rather than */zd/ or */md/ (which do not occur between vowels in adult English). Perhaps these errors represent the avoidance of word-final clusters such as /st/ or /nd/, provided that the resulting medial cluster is legal.

3. For all errors discussed here, *both plurals and past tense forms were affected*.

This demonstrates that both morphological affixes (plural -*s* and past tense -*ed*) must be accounted for within the same system,

as in standard linguistic models, and in Plunkett & Juola's (1999) connectionist simulation. McClelland & Rumelhart (1986), with a special network to generate past tense forms, cannot account for these data.

4. Note that verbs ending in /s/ were correct with [st] at 3;2, while verbs ending in /z/ were incorrect with [zəd].
Verbs ending in /s/ and /z/ are equally frequent in adult English (Francis & Kučera, 1982), which might lead to the expectation that they will be mastered at about the same time. However, [st] additionally occurs in hundreds of morphologically simple words

(e.g. *ghost, nest*) and has a much higher phonological frequency. The high frequency of [st] leads to past tense forms like *kissed* being mastered before forms like *used*.

This implies that past tense forms are produced in the same network as morphologically simple words,

as in standard linguistic models, but contra to most implementations of connectionist models and to Pinker & Prince (1988).

Connectionist Approaches

At 3;2, all the error types had in common that the syllabic allomorph was used.
The suffixal consonants [z ~ s] (plural) and [d ~ t] (past tense) were correctly produced, but with a schwa incorrectly added.
In words like *used* and *ghosts*, this was the only error.

At 3;3, connection weights changed to make it less likely that [ə] would be produced. **All three error types were correctly suppressed.**
This overgeneralized to forms that *should* have had [ə], **leading to affix-checking errors**.
McClelland & Rumelhart (1986:252) report that connectionist models are prone to this sort of error.

At 3;5, weights were changed to make the [ə] a more likely output.
Affix-checking errors disappeared, but other errors re-appeared.
(Except that [zd] did not regress. It was now a stable consonant cluster; other clusters also developed at this time.)

Connectionist models can adequately deal with the instance of U-shaped learning examined here.

Optimality Theory

relevant constraints (from Bernhardt & Stemberger1998):

NotTwice$_{\textbf{Dependents}}$(C-Root): **prevents** a sequence of two identical consonants: **horss*
motivates insertion of schwa **or** affix-checking

NotTwice(+grooved): **prevents** two adjacent high-amplitude consonants: **dishs*

NotTwice$_{Coda}$(Coronal,+voiced,-sonorant):
prevents two voiced obstruents in a coda: **zd*

Not(V-Root): **prevents** vowels from appearing in the output
here, it prevents the **insertion** of a [ə]
other versions of OT use: **DEP-V** or **FILL^{NUC}**

Distinct: two distinct elements in the input must be distinct in the output
prevents the /s/ of horse from also marking plural
prevents the /t/ of lift from also marking past tense

SinglyExpressed: a given piece of lexical information is expressed once
prevents doubling of suffixes in double-marking errors
prevents consonants and vowels from being produced twice in a word

Binary(Foot,ó): a foot must have two syllables
motivates augmentation of a monosyllabic word
doubling of the -s or -ed affix can achieve this augmentation

RE: **Which clusters allow insertion of schwa in plural and past tense?**

Bernhardt & Stemberger report that epenthesis in codas is rare in morphologically simple words: consonants usually delete from complex codas;

e.g. *fox* [fak] **or** [fas], **but not** *[fakəz].

Even in language development, schwa is not inserted for plural or past tense forms except between two coronal obstruents or between two fricatives, and then *only if the cluster is impossible* in the system at the time.

There is a set of constraints that prevents a vowel from being inserted between other consonants, e.g. between /l/ & /z/ in trolls *[troːləz].

We will not illustrate this in detail here, but simply assume it in the tables.

Examples of effects of ranking differences:

Distinct » Not(V-Root): pl. horses

insert [ə] to keep -*s* suffix & final /s/ of horse distinct

Not(V-Root) » Distinct: pl. horse

[s] corresponds to BOTH affix AND final /s/ of horse, which is preferable to inserting a vowel to keep them apart and distinct

Not(V-Root) » Binary(Foot,ó) » SinglyExpressed: pl. trolls

Plural -*s* is not doubled, even though the word has only one syllable

Binary(Foot,ó) » Not(V-Root) » SinglyExpressed: pl. trollses

Plural -*s* is doubled, so that the word can have two syllables

3;2 System at Point 1

The ranking of **Not(V-Root)** is unstable. It can be ranked high (preventing schwa):

	horse	horss	horses	dishs	dishes	juːzd	juːzəd	trolls	trollses
NotTwice(+grooved) NotTwice(C-Root)		*!		*!					
Distinct	*!								
Not(V-Root)			*		*		*!		*!
NotTwice(Cor,+vce,-son)						*			
Binary(Foot,ó)	*	*		*		*		*	
SinglyExpressed									*

NOTES:

1) Possible pronunciations are listed across the top. Relevant constraints are listed at the left.

2) An asterisk in a cell represents that the pronunciation in that column violates the constraint in that row. The exclamation point represents a fatal violation. Violations in shaded cells are unimportant.

3) The actual pronunciation is marked with a double-lined border.

Or **Not(V-Root)** can be ranked low, allowing schwa in various guises:

	horse	horss	horses	dishs	dishes	juːzd	juːzəd	trolls	trollses
NotTwice(+grooved) NotTwice(C-Root)		*!		*!					
Distinct	*!								
NotTwice(Cor,+vce,-son)						*!			
Binary(Foot,ó)	*	*		*		*		*!	
Not(V-Root)			*		*		*!		*!
SinglyExpressed									*

This ranking leads to errors like [juːzəd] being more common than errors like <u>trollses</u>, as was the case.

adult ranking: greater separation is needed between **Not(V-Root)** and the lesser-ranked constraints, so no violations even variably.

3;3 System at Point 2 (Regression I)

Not(V-Root) promoted (re-ranked higher).
This increased the separation from lower-ranked constraints, eliminating the associated errors.
It also variably raised the ranking higher than **Distinct**, leading to affix-checking errors:

	horse	horss	horses	dishs	dishes	juːzd	juːzəd	trolls	trollses
NotTwice(+grooved) NotTwice(C-Root)		*!		*!					
Not(V-Root)			*!		*		*!		*!
Distinct	*							(*)	
NotTwice(Cor,+vce,-son)						*			
Binary	*	*		*		*		*	
SinglyExpressed									*

The **promotion** of constraints is allowed by two approaches within OT:
Bernhardt & Stemberger (1998)
Boersma & Hayes (1999)
The necessary changes **cannot** be achieved in a third approach to learning in OT:
Tesar & Smolensky (2000)
The only possible re-ranking is downward: **Constraint Demotion**.
Because **Distinct** is ranked higher than **Not(V-Root)** at 3;2, *which is the ranking required by the adult system*, it cannot be demoted.
Instead, **NotTwice(Cor,+vce,-son)** and **Binary(Foot,ó)** would be demoted.
No regressions are predicted.

3;5 System at Point 3 (Regression II)

Since **Not(V-Root)** is ranked too high with respect to **Distinct**, it can be demoted *in all three approaches to learning*.

Affix-checking disappears, *but other errors re-appear*.

In an unrelated development, **NotTwice$_{\mathbf{Coda}}$(Coronal,+voiced,-sonorant)** is ranked lower, allowing the cluster [zd] to become stable; *[juːzəd] is impossible because of high-ranked constraints governing possible clusters (*not shown*).

	horse	horss	horses	dishs	dishes	juːzd	juːzdəd	trolls	trollses
NotTwice(+grooved) NotTwice(C-Root)		*!		*!					
Distinct	*!								
Binary	*	*		*		*!		*!	
Not(V-Root)			*!		*		*		*
SinglyExpressed									*
NotTwice(Cor,+vce,-son)						*			

CONCLUSIONS

1) U-shaped learning shows that constraints can be **PROMOTED**.

2) Tesar & Smolensky's limitations on change in the system during learning are inadequate to account for U-shaped learning.

3) Boersma & Hayes's approach works better (but see Stemberger, Bernhardt, & Johnson, 1999).

4) Models in which more kinds of change are possible work best, and can deal with changes that are correlated with changes in the frequency of an element or sequence in the system.

   - Bernhardt & Stemberger within OT

   - connectionist models

5) In order to account for interactions, there must be a single system for producing
   - plurals
   - past tense forms
   - morphologically simple words

Acknowledgment: We would like to thank Morgan Stemberger for her outstanding contribution to this research.

References

Berko, J. (1958). The child's learning of English morphology. *Word*, **14**, 150-177.

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

Boersma, P., & Hayes, B. (1999). Empirical tests of the Gradual Learning Algorithm. **Rutgers Optimality Archive ROA-348.**

Bybee, J.L., & Slobin, D.I. (1982). Rules and schemas in the development and use of the English past tense. *Language*, **58**, 265-289.

Francis, W.N., & Kucera, H. (1982). *Frequency analysis of English usage: lexicon and grammar*. Boston: Houghton Mifflin.

MacWhinney, B. (1978). The acquisition of morphophonology. *Monographs of the Society for Research in Child Development*, **43**, whole no. 1.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a Parallel Distributed Processing model of language acquisition. *Cognition*, **28**, 73-194.

Plunkett, K, & Juola, P. (1999). A connectionist model of English Past Tense and Plural Morphology. *Cognitive Science*, **23**, .

Rumelhart, D., & McClelland, J. (1986). On learning the past tenses of English verbs. In D. Rumelhart & J. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*, vol. 2 (pp216-271). Cambridge, MA: Bransford Books.

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

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