

Explaining Constraint Demotion in a Developing System

Suzanne Curtin and Kie Zuraw
University of Pittsburgh and University of Southern California

The task of the language learner in Optimality Theory is to identify the underlying forms and the constraint ranking that result in the correct output forms for the target language. This paper addresses the second point, constraint ranking. To understand how the learner arrives at a final ranking, we wish to know what is the initial ranking of constraints, and how the re-ranking of constraints proceeds. In this paper, we demonstrate how the gradual climb of a faithfulness constraint in the ranking accounts for the learner's variable and stage-like behavior, using truncation in Dutch as a case study.

Theories of learning should model the course of development—not just the endpoint—but little research in OT learnability has done so (but see Boersma & Levelt 1999 for acquisition of syllable shape). Variants of the Constraint Demotion Algorithm (CDA) (Tesar & Smolensky 1993, 1998, 2000; Prince & Tesar 2000; Hayes 1999) achieve the correct final ranking, but do not model the overlapping, non-discrete stages of acquisition, nor explain why one markedness constraint is demoted before another. We argue, however, that Boersma's (1998) Gradual Learning Algorithm (GLA) does capture the stage-like progression and variation found in the child language data, because of its gradient constraint ranking and sensitivity to lexical frequencies.

We first introduce Dutch children's prosodic development, then review the CDA and the GLA. Next, we present simulations using the GLA. Finally, we provide further support for the initial ranking of markedness >> faithfulness.

1. Stages of Development in Child Language

Children's early productions of polysyllabic words are initially truncated. Children start with unmarked structures (CV(C)), and gradually build up more marked representations (Demuth, 1995, 1998). This has been observed to follow a progression through four stages in Dutch (Fikkert, 1994), as illustrated in (1).

(1) Stages of prosodic development for 'balloon' [balɔn] (Fikkert 1994)

Stage 1 'one σ'	σ	lɔn
Stage 2 'one foot'	(σσ)	('balɔn)
Stage 3 'two feet'	(σσ)(σσ)	('ba)('lɔn)
Stage 4 'adult-like'	σ(σσ)	ba('lɔn)

* This work was supported by a SSHRC (Canada) doctoral grant #752-98-0283 awarded to S. Curtin.

Initially, the child's production is limited to a syllable; at Stage 2, the restriction is weakened to a foot. Stage 3 allows two equally stressed feet. Stage 4 permits a variety of structures.

The stages of prosodic development have been described as discrete (Fikkert, 1994; Demuth, 1997; Curtin, 1999). While this is a useful abstraction, in fact the stages overlap: at a given point in development, children may produce forms consistent with more than one stage, as in (2); children may also continue to produce forms consistent with an earlier stage, as illustrated in (3).

(2)	<i>Robin</i>		<i>Adult Form</i>	<i>Child Form</i>	<i>Gloss</i>	<i>Age</i>
	Stage 3	2 feet	'o:li;fant	['(o:li:) '(fant)]	'elephant'	2;4.8
	Stage 4	adult-like	'o:li;fant	['(o:li:) ,(ant)]	'elephant'	2;4.8
(3)	<i>Robin</i>		<i>Adult Form</i>	<i>Child Form</i>	<i>Gloss</i>	<i>Age</i>
	Stage 3	2 feet	spe:lo:'te:k	['(pɪlə)'(ti:k)]	'playground'	2;0.18
	Stage 2	1 foot	bə'lɒn	['(bu:ɒn)]	'balloon'	2;1.17

2. Learning algorithms

Several learning algorithms proposed in the OT acquisition literature use Constraint Demotion (Tesar & Smolensky, 1993, 1998; Prince & Tesar, 2000; Hayes, 1999). These algorithms are able to achieve the correct final ranking, but do not explain the overlapping, non-discrete stages of acquisition. We demonstrate that an algorithm using gradual re-ranking and probabilistic constraints (Boersma, 1998) is able to account for the stage-like progression and variation found in the child truncation patterns.

2.1 Constraint Demotion

The Constraint Demotion algorithm (CDA, Tesar & Smolensky 1993, 1998, 2000) serves as a model of the language learner. Its task is to determine the language specific constraint ranking based on the 'output' (surface) forms of the language. Initially, all the constraints are unranked. The learner compares the attested output to various other, sub-optimal candidates to identify the necessary dominating constraint(s): any constraints that are violated in the optimal output (and not in some rival output) must be dominated by some other constraint that rules out the rival. Constraints violated by the attested form are minimally demoted to immediately below the highest-ranking constraint that would rule out the sub-optimal candidate. A partial example is shown in (4). The output, candidate *a*, violates constraints C_2 and C_3 ; the rival, candidate *b*, violates C_4 . From the initial hierarchy of unranked $\{C_1, C_2, C_3, C_4\}$, C_2 is demoted below C_4 to yield $\{C_1, C_3, C_4\} \gg \{C_2\}$, and C_3 is demoted below C_4 to yield the stratified hierarchy $\{C_1, C_4\} \gg \{C_2, C_3\}$. The re-ranking of constraints occurs only in response to *positive evidence*—that is, constraint violation in the optimal form.

(4)

	C ₁	C ₂	C ₃	C ₄
a		*	*	
b				*

Most versions of the CDA extract all useful information from an input-output pair at the first pass. Because the transition from to the adult grammar is nearly instantaneous, the algorithm does not model learning stages.

2.2 Gradual Learning Algorithm

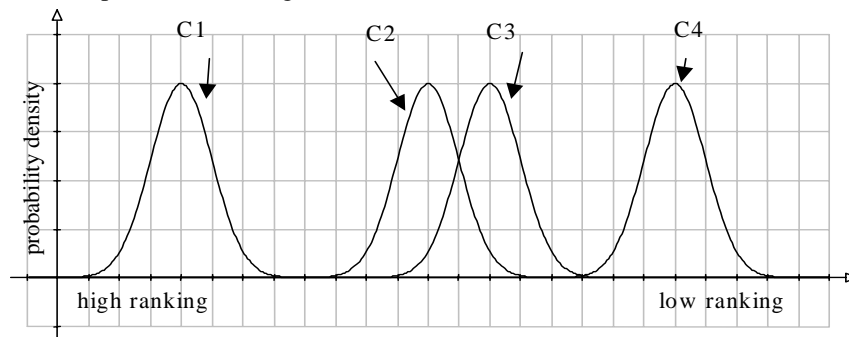
The GLA (Boersma 1998) is an elaboration of the CDA. Both algorithms are error-driven and shift constraint rankings in response to mismatches between adult forms and the current grammar's output. The GLA, however, uses stochastic constraint rankings: the ranking of constraints is neither fixed nor freely variable, but probabilistic (see also Hayes & MacEachern 1998). Each constraint in the grammar has a *ranking value* in arbitrary units. In any given utterance, the speaker generates an *effective value* for each constraint, as in (5), where *noise* is a constant, and *z* is a Gaussian random variable with mean 0 and standard deviation 1. For example, a constraint with a ranking value of 100 might be assigned an effective value of 100.5 on one occasion, 98.7 on another.

$$(5) \text{ effectiveValue} = \text{rankingValue} + \text{noise} * z$$

The constraints are then ordered by effective value to determine the constraint ranking to be used for that utterance. Each constraint can be thought of as associated with a probability density function centered on the ranking value, telling us how likely a given effective value is for that constraint (the closer to the ranking value, the more likely).

The figure in (6) illustrates a grammar with constraint C₁ nearly always top-ranked (high ranking values are on the left, as in an OT tableau), and C₄ bottom-ranked, but C₂ and C₃ variably ranked, with a preference for C₂ >> C₃.

(6) Example of stochastic grammar



When ranking values are far apart (like those of C_1 and C_2), the constraint ranking is effectively fixed: inversions of the usual ranking will be vanishingly rare. However, since the probability density functions' tails approach zero but never quite reach it, there is a very slight possibility that a ranking value might be selected at a position farther away from the center of the curve. For example, with a ranking spreading of 2, if C_i 's ranking value is 10 units greater than C_j 's, the chance of $C_j \gg C_i$ is one in 5000 (rare enough to be considered a speech error). When ranking values are close together, noticeable variation results.

The use of stochastic ranking allows Boersma's algorithm to model learning as gradual; ranking values are adjusted only slightly on each learning trial. In each learning trial, the learner compares its productions to the adult target. If there is a mismatch, the learner slightly lowers the rankings of all constraints that are violated more by the target than by the learner's form, and slightly raises the rankings of all constraints that are violated more by the learner's form more than by the target (the amount of adjustment, or "plasticity", is pre-set).

3. The simulation

We used the progression from truncation to non-truncation in Dutch as a test case for the GLA. Our goal was to model the progression of early truncation patterns and the variable/overlapping productions seen in the Dutch child language. We did not seek to model the acquisition of stress placement, which occurs later and requires additional constraints.

3.1 Learning data and constraints

The algorithm was provided with adult output forms for Dutch polysyllabic words of up to four syllables, with all possible locations of primary stress (acute accent and bold in (7)). Type frequencies for each prosodic pattern were also provided (CELEX database¹).

(7) <i>Pattern</i>	<i>Frequency</i>
σ σ	1622
σ σ	681
σσ σ	184
σ σσ	360
σσ σ	284
σσσσ	91
σ σσσ	126
σσ σσ	132
σσσσ σ	114

Because truncation, not stress placement, was the object of the simulation, the inputs omit segmental and featural information. As a simplification, secondary stress was not encoded. The constraints used in the simulation include

1. http://europa.eu.int/celex/htm/celex_en.htm

markedness constraints that encourage truncation (8), and faithfulness constraints that resist truncation (9).

- (8) ALLσLEFT (McCarthy & Prince 1993; Ito & Mester 1997; Spaelti 1997)
Align all syllables to the left edge of the Prosodic Word. Favors single-syllable outputs.

ALLFTLEFT (McCarthy & Prince 1993; Ito & Mester 1997; Spaelti 1997)
Align all feet to the left edge of the Prosodic Word (one violation for each misaligned foot). Favors maximally one-foot outputs.

FOOTMAX
Feet are maximally disyllabic.

PARSEσ (Prince & Smolensky 1993)
All syllables must be parsed into feet. With FTMAX and ALLFTLEFT, favors truncation in 3- and 4-syllable words (FTMAX prohibits one big foot, ALLFTL prohibits multiple feet, and PARSEσ prohibits unfooted syllables).

- (9) MAXPITCHPROM (Curtin 1999)
A prominent vowel in the input has a correspondent in the output (MAXPITCH). Favors retaining the stressed syllable of the adult form.

MAXFINALPROM (Curtin 1999)
A final syllable in the input has a correspondent in the output (MAXFINAL). Favors retaining the final syllable of the adult form.

MAXσ
A sonority peak in the input has a nuclear correspondent in the output. Disfavors any truncation.

As an example, (10) shows constraint violations for input / $\sigma_1\sigma_2\sigma_3$ /

(10)

/ $\sigma_1\sigma_2\sigma_3$ /	ALLσLT	ALLFTL	FOOT MAX	PARSEσ	MAXσ	MAX PITCH	MAX FINAL
($\sigma_1\sigma_2\sigma_3$)	***		*				
($\sigma_1\sigma_2$)(σ_3)	***	*					
(σ_1)($\sigma_2\sigma_3$)	***	*					
($\sigma_1\sigma_2$) σ_3				*			
σ_1 ($\sigma_2\sigma_3$)		*		*			
($\sigma_1\sigma_2$)	*				*		*
($\sigma_1\sigma_3$)	*				*		
($\sigma_2\sigma_3$)	*				*	*	
(σ_1)					**		*
(σ_2)					**	*	*
(σ_3)					**	*	*

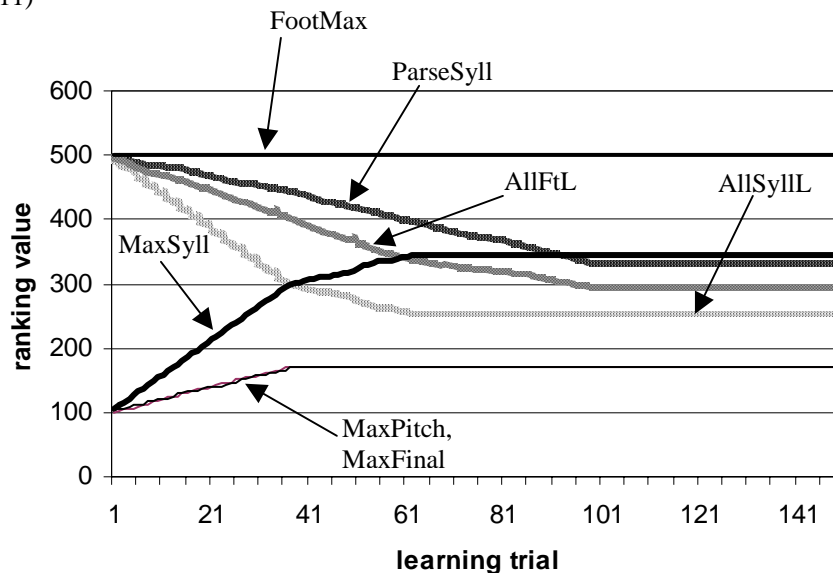
3.2 Simulation Details

Using OTSoft (Hayes, Tesar & Zuraw 2000), initial plasticity was set at 0.6; final plasticity at 0.006 (the software interpolates). Because children's early productions are less marked than the target adult productions, it is often proposed that children begin with markedness constraints outranking faithfulness constraints (Demuth 1995, Gnanadesikan 1995, Levelt 1995, Smolensky 1996). Therefore, we set the initial ranking values at 500 for all markedness constraints, and 100 for all faithfulness constraints. We ran 420 learning trials, where each trial examines 9 forms, chosen at random but proportionately to the frequencies given in (7). Thus, there were many more learning trials for adult $\sigma\sigma$ than for $\sigma\sigma\sigma\sigma$.

3.3 End Result of Simulation

The constraints' ranking values over time are shown in **Error! Reference source not found.** As expected, the learning algorithm reached a final stage that corresponded to adult (untruncated) productions. At the end of learning, MAX- σ is decisively top-ranked, preventing truncation.

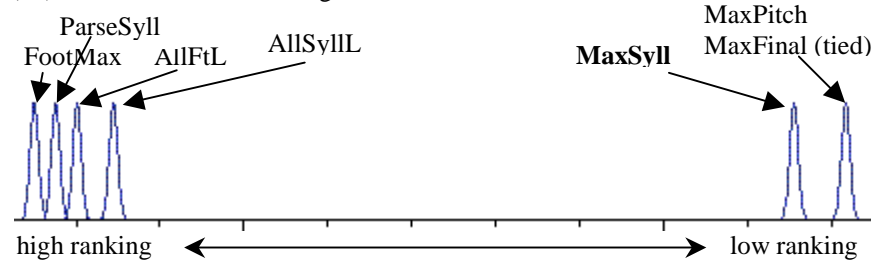
(11)



3.4 Stages and Variation

We examined the learner's grammar at several key points during the simulation, to determine what truncation patterns would result. After 7 learning trials, the grammar is as shown in (12).

(12) Grammar after 7 learning trials



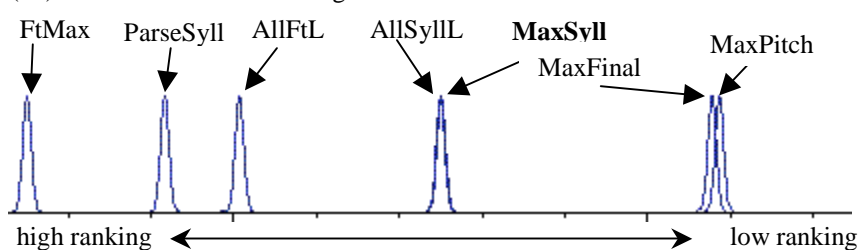
We tested the behavior of this grammar by running the target forms through it 2000 times. Each such test trials uses slightly different effective values, generated from the same ranking values. The results are shown in (13). The first column, for example, shows that for an input like /bá.di/, the grammar truncates to [ba] 50% of the time, and to [di] 50% of the time. Because ALL σ LT decisively outranks MAX σ , outputs are monosyllabic. Tied MAXPITCH and MAXFINAL determine which syllable to retain—the adult pitch-prominent, or the adult final syllable. This behavior corresponds to the Stage 1, truncation to a monosyllable.

(13) Output frequencies after 7 learning trials

	/ σ /	/ σ /	/ $\sigma\sigma$ /	/ $\sigma\sigma$ /	/ $\sigma\sigma$ /	/ $\sigma\sigma\sigma$ /	/ $\sigma\sigma\sigma$ /	/ $\sigma\sigma\sigma$ /	/ $\sigma\sigma\sigma$ /
(σ_1)	50%		50%			50%			
(σ_2)	50%	100%		50%			50%		
(σ_3)			50%	50%	100%			50%	
(σ_4)						50%	50%	50%	100%

After 37 learning trials, the grammar is as shown in (14), and the outputs as in (15). ALL σ LT is the first markedness constraint to be noticeably demoted, because all inputs (2-, 3-, or 4-syllable) encourage its demotion. ALL σ LT is still ranked high enough to force some monosyllabic outputs, but MAX σ is now ranked almost as high to allow some disyllabic outputs. High-ranking ALLFTL and FOOTMAX rule out fully footed 3- and 4-syllable productions, but PARSE σ rules out unfooted syllables; thus, productions longer than two syllables are not allowed. This corresponds to a mixture of the Stage 1 and Stage 2.

(14) Grammar after 37 learning trials

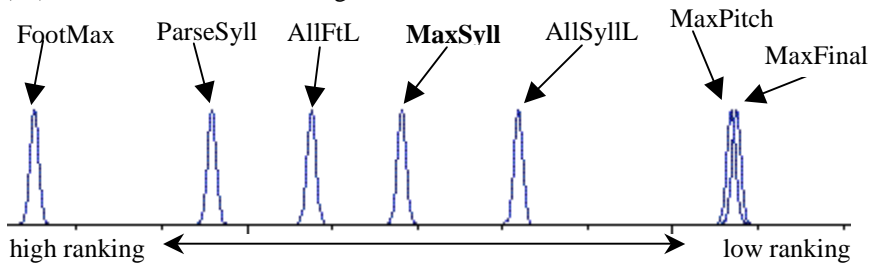


(15) Output frequencies after 37 learning trials

	/σ/	/σσ/	/σσσ/	/σσσσ/	/σσσσσ/	/σσσσσσ/	/σσσσσσσ/	/σσσσσσσσ/	/σσσσσσσσσ/
(σ ₁)	6%		6%			6%			
(σ ₂)	49%	55%		6%			6%		
(σ ₃)			49%	49%	55%			6%	
(σ ₄)						49%	49%	49%	55%
(σ ₁ σ ₂)	45%	45%							
(σ ₁ σ ₃)			45%		23%				
(σ ₁ σ ₄)						45%			15%
(σ ₂ σ ₃)				45%	23%				
(σ ₂ σ ₄)							45%		15%
(σ ₃ σ ₄)								45%	15%

The grammar after 50 learning trials is shown in (16), and the grammar's outputs are shown in (17). MAX σ now consistently outranks ALL σ LT, so truncation to a monosyllable no longer occurs. Because ALLFTL, FOOTMAX, and PARSE σ still high ranked, productions longer than two syllables are still not possible, and we see truncation to a disyllable. The syllables retained are the pitch-prominent syllable and the final syllable; if they are the same, the choice of which other syllable to retain. This corresponds to the child's Stage 2 (truncation to a disyllable).

(16) Grammar after 50 Learning Trials



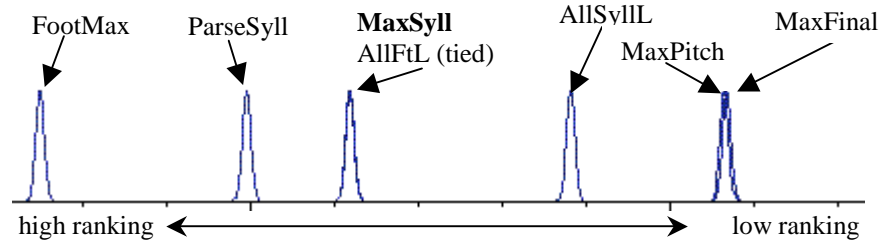
(17) Outputs after 50 learning trials

	/σ/	/σσ/	/σσσ/	/σσσσ/	/σσσσσ/	/σσσσσσ/	/σσσσσσσ/	/σσσσσσσσ/	/σσσσσσσσσ/
(σ ₁ σ ₂)	100%	100%							
(σ ₁ σ ₃)			100%	50%					
(σ ₁ σ ₄)					100%				33%
(σ ₂ σ ₃)				100%	50%				
(σ ₂ σ ₄)						100%			33%
(σ ₃ σ ₄)							100%		33%

The grammar after 62 learning trials is shown in (18), and its outputs are shown in (19). ALL σ LT is the next markedness constraint to be demoted. This is because evidence for its low ranking comes only from 3- and 4-syllable words. MAX σ and ALLFTL are now tied in ranking value, so productions with more

than one foot have begun to occur. Because PARSE σ is high-ranked, all syllables are footed. This behavior corresponds to variation between the child's Stage 2 (truncation to a disyllable) and Stage 3 (no truncation, no unfooted syllables).

(18) Grammar after 62 learning trials

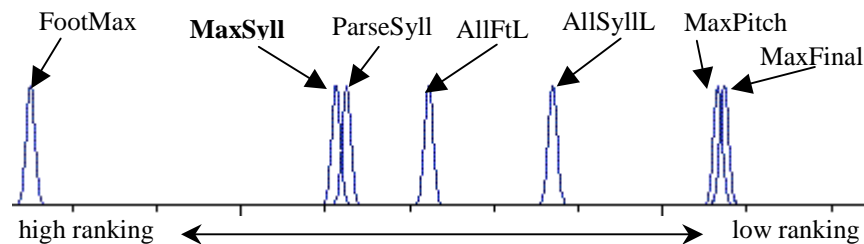


(19) Outputs after 62 learning trials

	/ $\acute{\sigma}$ /	/ $\sigma\acute{\sigma}$ /	/ $\acute{\sigma}\sigma\sigma$ /	/ $\sigma\acute{\sigma}\sigma$ /	/ $\acute{\sigma}\sigma\sigma\sigma$ /	/ $\sigma\acute{\sigma}\sigma\sigma$ /	/ $\sigma\sigma\acute{\sigma}\sigma$ /	/ $\sigma\sigma\sigma\acute{\sigma}$ /
$(\sigma_1\sigma_2)$	100%	100%						
$(\sigma_1\sigma_3)$			48%					
$(\sigma_1\sigma_4)$				24%	48%			16%
$(\sigma_2\sigma_3)$				52%	24%			
$(\sigma_2\sigma_4)$						48%		16%
$(\sigma_3\sigma_4)$							48%	16%
$(\sigma_1)(\sigma_2\sigma_3)$			52%	48%				
$(\sigma_1\sigma_2)(\sigma_3)$				52%				
$(\sigma_1\sigma_2)(\sigma_3\sigma_4)$					52%	52%	52%	52%

Finally, after 94 learning trials, truncation ceases, under the grammar shown in (20). MAX σ now outranks all competing markedness constraints, preventing truncation. Because PARSE σ outranks ALLFTL, all syllables are footed. This corresponds to the child's Stage 3. Other constraints, including weight-to-stress, anti-clash, and perhaps faithfulness to lexical specifications, would be necessary to attain the adult stress system. PARSE σ is demoted last, because only certain stress patterns give evidence for its low ranking.

(20) Grammar after 94 learning trials—final behavior achieved



The order in which the markedness constraints are demoted below MAX σ (ALL σ L, then ALLFTL, then PARSE σ) is important. If they had been demoted in

a different order, the learner's behavior would not necessarily resemble the attested child stages. For example, if PARSE σ were demoted below ALLFTL, we see an odd stage between Stage 2 and Stage 3, with no truncation, but only one foot per word (ALLFTL >> MAX σ >> {PARSE σ , ALL σ L}). And if ALLFTL were demoted before ALL σ L, there would be no stage of truncation to a disyllable.

3.5 The initial state

The next simulation explored the question of whether it is necessary for the initial state to have markedness constraints ranked above faithfulness constraints. To test this, we ran the algorithm on the same set of input and output candidates with the same constraints.

The parameters were kept the same as in the initial simulation. However, the initial ranking of markedness and faithfulness was changed: all constraints were initially assigned a ranking value of 100.

The result is that the initial grammar produces free variation among nontruncation, partial truncation, and extreme truncation, as shown in (21).

(21) Outputs when all constraints' ranking value = 100

	/ σ /	/ $\sigma\sigma$ /	/ $\sigma\sigma\sigma$ /	/ $\sigma\sigma\sigma\sigma$ /	/ $\sigma\sigma\sigma\sigma\sigma$ /	/ $\sigma\sigma\sigma\sigma\sigma\sigma$ /	/ $\sigma\sigma\sigma\sigma\sigma\sigma\sigma$ /	/ $\sigma\sigma\sigma\sigma\sigma\sigma\sigma\sigma$ /	/ $\sigma\sigma\sigma\sigma\sigma\sigma\sigma\sigma\sigma$ /
σ_1	21%	0	21%	0	0	21%	0	0	0
σ_2	21%	50%	0	21%	0	0	21%	0	0
σ_3	0	0	21%	21%	50%	0	0	21%	0
σ_4	0	0	0	0	0	21%	21%	21%	50%
$\sigma_1\sigma_2$	58%	50%	0	0	0	0	0	0	0
$\sigma_1\sigma_3$	0	0	33%	0	13%	0	0	0	0
$\sigma_1\sigma_4$	0	0	0	0	0	17%	0	0	6%
$\sigma_2\sigma_3$	0	0	0	33%	13%	0	0	0	0
$\sigma_2\sigma_4$	0	0	0	0	0	0	17%	0	5%
$\sigma_3\sigma_4$	0	0	0	0	0	0	0	17%	6%
$\sigma_1\sigma_2\sigma_3$	0	0	25%	25%	25%	0	0	0	0
$\sigma_1\sigma_2\sigma_4$	0	0	0	0	0	4%	4%	0	0
$\sigma_1\sigma_3\sigma_4$	0	0	0	0	0	4%	0	4%	0
$\sigma_2\sigma_3\sigma_4$	0	0	0	0	0	0	4%	4%	0
$\sigma_1\sigma_2\sigma_3\sigma_4$	0	0	0	0	0	33%	33%	33%	33%

The more truncated forms will gradually become less frequent as MAX- σ is promoted. But in the child data, the progression is not from wide variation to narrower variation to consistent non-truncation. Rather, it is from extreme truncation to non-truncation, passing through variable intermediate stages. Beginning with the markedness constraints top-ranked is the only way to capture the earliest stages, in which extreme truncation is consistently observed.

The results of this simulation thus provide support for the initial ranking of markedness above faithfulness.

4. Conclusion

The results of our simulation demonstrate that a probabilistic, gradual learning procedure (specifically, Boersma's GLA) can mimic the developmental behavior of children acquiring polysyllabic words, as long as markedness constraints initially outrank faithfulness constraints. The probabilistic nature of the grammar accounts for variability, and the gradual nature of the learning procedure produces a stage-like progression. The initial ranking of markedness constraints over faithfulness constraints accounts for the child's progression from less marked to more marked structure, rather than merely from variation to stability.

References

- Boersma, Paul (1998). *Functional phonology*. Doctoral dissertation, University of Amsterdam. The Hague: Holland Academic Graphics.
- Boersma, Paul & Levelt, Clara (1999). Gradual constraint ranking learning algorithm predicts acquisition order. *Proceedings of the 30th Child Language Research Forum*, to appear.
- Curtin, Suzanne (1999). Positional prominence in the acquisition of prosodic structure. *Proceedings of the Annual Meeting of the Canadian Linguistics Society*, (pp. 89-100).
- Curtin, Suzanne (2001). Children's Early Representations: Evidence from Production and Perception. To appear in the Proceedings of the Holland Institute of Linguistics Phonology Conference (HILP 5), Potsdam, Germany.
- Curtin, Suzanne (2001). *Representational Richness in Phonological Development*. Doctoral dissertation, Department of Linguistics, University of Southern California.
- Demuth, Katherine (1995). Markedness and the development of prosodic structure. In *NELS 25*, GLSA, University of Massachusetts, Amherst.
- Fikkert, Paula (1994). *On the acquisition of prosodic structure*. Holland Institute of Generative Linguistics.
- Gnanadesikan, Amalia (1995). Markedness and faithfulness constraints in child phonology. Ms., University of Massachusetts, Amherst. (ROA-67).
- Hayes, Bruce (1999). Phonological acquisition in Optimality Theory: The early stages. Ms., Linguistics Department, Los Angeles. ROA-327-0699.
- Hayes, Bruce & MacEachern, Margaret (1998). Quatrain Form in English Folk Verse. *Language*, 74, 473-507.
- Hayes, Bruce, Tesar, Bruce, & Zuraw, Kie. (2000). OTSoft, software package, <http://www.linguistics.ucla.edu/people/hayes/otsoft/>
- Itô, Junko & Mester, Armin (1997). Sympathy and German truncations. In V. Miglio & B. Moren (Eds.), *Proceedings of the Hopkins Optimality Workshop/Maryland Mayfest*. University of Maryland Working Papers in Linguistics, 5.
- Levelt, Clara (1995). Unfaithful kids: Place of articulation patterns in early child language. Department of Cognitive Science, Johns Hopkins University, Baltimore, MD. Talk.
- McCarthy, John & Prince, Alan (1993). Generalized alignment. In G. Booij & J. van

- Marley (Eds.), *Yearbook of Morphology*, 79-153. Dordrecht: Kluwer.
- McCarthy, John & Prince, Alan (1995). Faithfulness and reduplicative identity. In J. Beckman, L. Walsh Dickey & S. Urbanczyk (Eds.), *University of Massachusetts Occasional Papers in Linguistics 18: Papers in Optimality Theory*, (pp. 249-384). Amherst: GLSA.
- Prince, Alan & Tesar, Bruce (2000). Learning phonotactic distributions. Ms., Linguistics Department, Rutgers University, New Brunswick, NJ. ROA-353-1099.
- Spaelti, Philip (1997). *Dimensions of variation in multi-pattern reduplication*. Doctoral dissertation, Department of Linguistics, University of California, Santa Cruz.
- Smolensky, Paul (1996). The initial state and 'richness of the base' on Optimality Theory. Technical Report JHU-CogSci-96-4, Cognitive Science Department, Johns Hopkins University, Baltimore, MD. ROA-154.
- Tesar, Bruce (2000). Using inconsistency detection to overcome structural ambiguity in language learning. Technical Report RcCCS-TR-58, Rutgers Center for Cognitive Science, Rutgers University, New Brunswick, NJ. Rutgers Optimality Archive ROA-426.
- Tesar, Bruce & Smolensky, Paul (1993). The learnability of Optimality Theory: An algorithm and some basic complexity results. Technical Report CU-CS-687-93, University of Colorado at Boulder, Boulder, CO. ROA-2.
- Tesar, Bruce & Smolensky, Paul (1998). Learnability in Optimality Theory. *Linguistic Inquiry*, 29, 229-268.
- Tesar, Bruce & Smolensky, Paul (2000). *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.
- Zuraw, Kie (2000). *Patterned exceptions in phonology*. Doctoral dissertation, Department of Linguistics, University of California, Los Angeles.