Weighted Constraints and Faithfulness Cumulativity in Phonological Acquisition

Ashley Farris-Trimble
University of Iowa

1. Introduction

Most acquisition research, at least that done on typically-developing populations, takes as a starting assumption the fact that the target grammar will eventually be acquired. The path of development is the more interesting focus. On the other hand, many simulations of learning have focused on whether the model can learn the target grammar, without attending to whether the learning process also models real-life acquisition. The goals of this paper are first, to introduce and account for a particular stage that can arise in phonological development and second, to use a learning simulation to characterize the conditions that bring about this acquisition stage.

The acquisition stage introduced here is the cumulative faithfulness effect (CFE; Farris-Trimble, 2008). CFEs occur when multiple individual unfaithful mappings are allowed, but those unfaithful mappings cannot combine in some domain. From a theoretical point of view, the focus here is on constraint-based accounts of phonological acquisition. For instance, optimality theory (Prince & Smolensky, 1993/2004) is a constraint-based account in which markedness constraints and faithfulness constraints are in a hierarchical ranking that determines the optimal output for any given input. CFEs arise when multiple markedness constraints outrank or outweigh multiple faithfulness constraints, which themselves are differentially ranked or weighted, forcing the grammar to choose between multiple unmarked (and thus unfaithful) mappings. First-language acquisition is a prime source for CFEs: in the early stages of acquisition, children's outputs tend to be very unfaithful, and so the potential for multiple unfaithful mappings in a single domain is increased. Most researchers agree that in the earliest stages of learning, markedness constraints outrank faithfulness constraints, thus setting up the type of grammar in which CFEs are likely to be found.

In the following sections, a CFE found in Amahl's acquisition of English (Smith, 1973) is described and accounted for. We then turn to the question of

* Thanks to Dan Dinnsen, Judith Gierut, Bob McMurray, Joe Toscano and the MAClab and Word Learning Lab at the University of Iowa, as well as participants at BUCLD 33. Supported by grants from the NIH to Indiana University: DC00012 and DC001694; and to the University of Iowa: DC000242.

how the CFE might have come about, using a learning algorithm to model both the CFE stage and the later correct acquisition of the target grammar.

2. The CFE

An example of a CFE in first-language acquisition comes from Smith's (1973) influential diary study of Amahl, a typically developing child learning British English. Smith details Amahl's phonological acquisition from about two to four years of age; the examples in (1) come from Amahl's data between ages 2;3 (years;months) and 2;5.[1] During this stage, Amahl exhibited an allophonic voicing pattern in the stop series: all word-initial stops were realized as voiceless unaspirated lenis, all word-medial stops were realized as voiced, and all word-final stops were realized as voiceless fortis. The most relevant of these patterns for the issue at hand is the word-final devoicing of all stops, as is shown in (1a). The second pattern is fricative stopping: coronal fricatives in all word positions are realized as stops, with the voicing corresponding to the allophonic pattern, as in (1b). The CFE arises when we examine the behavior of target word-final voiced fricatives. Given the evidenced stopping and devoicing patterns, we might expect word-final /z/ to be realized as [t]; instead, this segment is deleted entirely. This pattern is shown in (1c).

(1) Amahl (2;3 to 2;5)

a. Word-final stop devoicing

[b̥ɛt]	'bed'	[g̊uːp]	'cube'	[ɛk]	'egg'
[d̥ɛt]	'red'	[wʌp]	'rub'	[g̊ɛk]	'leg'

b. Fricative stopping

[b̥ʌt]	'bus'	[d̥ʌn]	'sun'	[d̥u]	'zoo'
[b̥ʌt]	'brush'	[d̥əːt]	'shirt'	[d̥ɛ]	'there'
[b̥aːt]	'bath'	[d̥əːd̥iː]	'thirsty'		

c. Word-final voiced fricatives delete

[nɔiː][2]	'noise'	[d̥iː]	'cheese'	[b̥iː]	'please'
[nuː]	'nose'	[g̊aːgiː]	'glasses'	[d̥idə]	'scissors'
[b̥aiə]	'pliers'	[nuː]	'news'		

---

[1] These are Smith's Stages 1-4 of Amahl's acquisition, which occurred between the 60th and 137th days of his second year.

[2] Amahl frequently, but not consistently, lengthens the vowel before an underlyingly voiced obstruent, whether it is deleted or devoiced. The lengthening of a vowel before a deleted obstruent produces a counterbleeding opacity effect. As counterbleeding is not the focus of this analysis, this effect will be ignored here. For two different methods for dealing with counterbleeding in optimality theory, see McCarthy (1999, 2007).

The constraints necessary to account for Amahl's CFE are listed in (2). The markedness constraints ban fricatives and voiced coda obstruents. Faithfulness constraints militating against change in voice or manner, as well as deletion, are also relevant.

(2) Constraints relevant for Amahl's CFE
*FRIC: Fricatives are banned
*VOICODA: Voiced obstruents are banned in coda position
IDENT[voice]: Input and output correspondents have the same value for the feature [voice]
IDENT[continuant]: Input and output correspondents have the same value for the feature [continuant]
MAX: Input segments have output correspondents

A ranking paradox, shown in (3), occurs in a standard OT account of Amahl's CFE. Because voiced obstruents never appear in coda position and fricatives never occur at all in Amahl's outputs, it is clear that the markedness constraints against them must be high-ranked. The ranking paradox occurs among the faithfulness constraints, however. MAX must be ranked above IDENT[voice] and IDENT[continuant] to compel preservation of the final obstruents in words like *bed* and *bus*, respectively, as in (3a,b). The high ranking of MAX, however, would eliminate the attested deletion candidate in (3c), [nɔiː] for input *noise*, in favor of the doubly-unfaithful output, [nɔit].

(3) Standard OT fails to account for Amahl's CFE
a. MAX >> IDENT[voice]

/bɛd/ 'bed'	*VOICODA	*FRIC	MAX	ID[voice]	ID[cont]
a. b̥ɛd	*!				
b. ☞ b̥ɛt				*	
c. b̥ɛ			*!		

b. MAX >> IDENT[continuant]

/bʌs/ 'bus'	*VOICODA	*FRIC	MAX	ID[voice]	ID[cont]
a. b̥ʌs		*!			
b. ☞ b̥ʌt					*
c. b̥ʌ			*!		

c. Ranking paradox

/nɔiz/‘noise’	*VOICODA	*FRIC	MAX	ID[voice]	ID[cont]
a. nɔiːz	*!	*			
b. nɔis		*!		*	
c. nɔiːd	*!				*
d. ☞ nɔit				*	*
e. ☜ nɔiː			*!		

In Amahl’s grammar, the change from input /s/ or /d/ to output [t] is allowed, but the change from /z/ to [t] is not. The problem with the change from target /z/ to output [t] is not one of markedness, as [t] is a completely unmarked segment and an acceptable substitute for /s/ and /d/. Rather, the trouble with substituting [t] for /z/ is that it is too unfaithful. Amahl’s grammar must choose between two equally unmarked outputs for a word like *noise*: [nɔit] and [nɔiː]. The output [nɔit] violates two relatively low-ranked faithfulness constraints, IDENT[voice] and IDENT[continuant], but the output [nɔiː] violates one relatively high-ranked constraint, MAX. Amahl’s grammar chooses the candidate that violates only a single faithfulness constraint, even though it is higher-ranked, but OT has no way of allowing this mapping.

Harmonic Grammar (HG; Legendre, Miyata & Smolensky, 1990a,b; Smolensky & Legendre, 2006) is an alternative constraint-based theory. HG was a precursor to OT and was originally intended to model connectionist networks. Each phonological input and output can be thought of as a node in the grammar, with links between them symbolizing input-output pairs. Each link has a weight; the cumulative weight of all the links between an input and an output determines its activation. If heavier weights are given to more likely outputs, or more likely input-output pairs, then the resulting candidates are more likely to be activated in the grammar. HG was originally rejected in favor of OT because HG was argued to predict some grammars that do not seem to occur in the linguistic typology. More recently, though, Pater, Bhatt, and Potts (2007) have shown that HG actually predicts a limited range of languages, particularly if restrictions are placed on the domain of evaluation of certain constraints, and HG has had a resurgence (e.g., Jesney & Tessier, 2007; Pater, Bhatt & Potts, 2007; Pater, Jesney & Tessier, 2007; Goldrick & Daland, in press).

HG differs from OT in that constraints are weighted rather than ranked. Constraints with greater weights would translate into higher-ranked constraints in OT, while low-weight constraints are similar to low-ranked constraints. Moreover, a candidate’s violations are summed across the entire constraint set, so that even low-weight constraints make a contribution to the candidate’s harmony (the sum of the candidate’s violations). The resulting crucial difference between the two models is that strict domination is a key feature of OT but not of HG. Because of the symbolic nature of OT's constraints, a higher-ranked constraint strictly dominates a lower-ranked one—no number of violations of the lower-ranked constraint can overcome the violation of a higher-ranked constraint (McCarthy, 2002). On the other hand, in HG, multiple violations of low-weight constraints may, when added together, "gang up" on a higher-weight constraint, allowing low-weight HG constraints to have more power than low-ranked constraints in OT.

Amahl's CFE can be obtained in an HG account by weighting the faithfulness constraints such that a single violation of MAX outweighs a single violation of either of the IDENT constraints, but the cumulative violation of both IDENT constraints outweighs MAX. In other words, the violations of IDENT[voice] and IDENT[continuant] trade off for a violation of MAX. Relevant weighting arguments are given in (4), where the ≻ symbol in the right-hand column indicates "more harmonic than."

(4) Constraint weightings necessary for Amahl

Weighting	Result	
$W_{*\text{VoiCoda}} > W_{\text{ID[voice]}}$	/bɛd/	bɛt ≻ bɛd
$W_{*\text{Fricative}} > W_{\text{ID[cont]}}$	/bʌs/	bʌt ≻ bʌs
$W_{\text{Max}} > W_{\text{ID[voice]}}$	/bɛd/	bɛt ≻ bɛ
$W_{\text{Max}} > W_{\text{ID[cont]}}$	/bʌs/	bʌt ≻ bʌ
$W_{\text{Max}} < W_{\text{ID[voice]}} + W_{\text{ID[cont]}}$	/nɔiz/	nɔi ≻ nɔit

For space considerations, the tableaux in (5) only show the faithfulness constraints and unfaithful candidates; as shown above, the marked candidates are eliminated by the high weight of the markedness constraints. The importance of the weight of MAX relative to each of the IDENT constraints is shown in (5a,b), where the weight of MAX is 1.5 and the weight of each of the IDENT constraints is 1. In (5c), however, the cumulative harmony of candidate d., the doubly-unfaithful candidate, is -2 (the sum of the weights of each of the IDENT constraints; note that because harmony values are negative, smaller absolute values correspond to higher harmony values). Candidate d. is less harmonic than candidate e., which incurs only a single violation of MAX and thus wins.

(5) HG account of Amahl's CFE

a. $W_{\text{Max}} > W_{\text{ID[voice]}}$

/bɛd/ 'bed'	MAX w=1.5	ID[voice] w=1	ID[cont] w=1	H
a. ☞ b̥ɛt		-1		-1
b. b̥ɛ	-1			-1.5

b. $W_{\textsc{Max}} > W_{ID[continuant]}$

/bʌs/ 'bus'	MAX w=1.5	ID[voice] w=1	ID[cont] w=1	H
a. ☞ b̥ʌt			-1	-1
b. b̥ʌ	-1			-1.5

c. $W_{\textsc{Max}} < W_{ID[voice]} + W_{ID[continuant]}$

/nɔiz/ 'noise'	MAX w=1.5	ID[voice] w=1	ID[cont] w=1	H
a. nɔit		-1	-1	-2
b. ☞ nɔiː	-1			-1.5

The next section turns to the question of how Amahl acquired his CFE.

3. The weight of MAX

One of the most important issues in dealing with any acquisition-related phenomenon is determining how the child arrived at his grammar. How did Amahl acquire a grammar in which multiple faithfulness violations were dispreferred? In this section, we examine Jesney and Tessier's (2007) adaptation of the Gradual Learning Algorithm to HG, henceforth HG-GLA. We use a simulated learner to show that HG-GLA predicts a stage at which CFEs will arise, while also predicting eventual acquisition of the target language.

The Gradual Learning Algorithm (GLA; Boersma, 1998) is a computational algorithm proposed to account for the gradual reranking of constraints during the course of acquisition. The GLA has been modified to account for the reweighting of constraints in HG (e.g., Jesney & Tessier, 2007; Pater, Bhatt & Potts, 2007; Pater, 2008). In the algorithm, the gradual change in constraint weights is a result of mismatch between the child's (incorrect) output and the adult target (the learning datum). The recognition of a mismatch spurs the grammar to slightly increase the weight of every constraint violated by the child's incorrect output, while also decreasing the weight of every constraint violated by the observed learning datum. The amount by which the weight of a constraint is increased or decreased is known as the plasticity. A small amount of noise is introduced into the algorithm at each learning step to allow for stochastic evaluation. In this way, constraint weights are changed by some small increment every time the child perceives a mismatch; over time, the weights become adult-like.

The HG-GLA requires a number of assumptions, both broad and narrow. Any error-driven learning algorithm must assume that the child can recognize a mismatch between her output and the adult target, and that the child knows the constraint relevant to the mismatch. More specifically, the computational instantiation of the HG-GLA used here assumes that outputs with equal weights occur in free variation, and that the learning data are essentially an even distribution of word types. Some other proposals are also considered here: first, that output-oriented (that is, markedness and output-to-output faithfulness) constraints have an initial-state weight that is greater than the weight of the input-output constraints, and second, that the plasticity of markedness constraints is greater than the plasticity of faithfulness constraints (Jesney & Tessier, 2007).

Amahl's learning was implemented in the HG-GLA using Praat (Boersma & Weenink, 2008). Markedness constraints were assigned an initial weight of 100[3] and faithfulness constraints an initial weight of 0. The plasticity was set at 0.1. The learning data were generated from an adult grammar such that the simulation received an equal number of exemplars of *bed*, *bus*, and *noise*. The algorithm was presented with 1200 learning trials, and the simulation was repeated in ten blocks. Results reported here are averaged across all ten blocks.

Figure 1 shows the trajectories of constraint weights over the 1200 learning trials. The solid line represents the two markedness constraints; because the grammar received the same number of words with fricatives as it did words with voiced codas, *FRICATIVE and *VOICODA follow the same trajectory, as do IDENT[cont] and IDENT[voice], which are represented by the larger dashed line. The small dotted line represents MAX, and the double line represents the cumulative weight of the two IDENT constraints. The markedness constraints begin with a weight of 100, while the faithfulness constraints begin with a weight of 0. In order to result in a CFE, it is necessary for the weight of each of the individual IDENT constraints to have a weight less than MAX, but for their sum to be greater than the weight of MAX. That is exactly what we see in Figure 1. Soon after the start of learning, the faithfulness constraints begin to diverge, with the weight of MAX increasing faster than that of either of the IDENT constraints but slower than their sum. Before long, MAX has a weight greater than either of the IDENT constraints, but smaller than the sum of the IDENT constraints. This is exactly the relative weighting necessary for a CFE. This instantiation of the HG-GLA goes through a CFE stage while resulting in an adultlike output at the final stages.

---

[3] Note that in §2, the HG tableaux have shown weightings in the small digits (e.g. 1, 2.5, etc.). The switch to larger numbers in this section (e.g., a range from 0-100) is simply a shift in perspective. In all cases, what is crucial is the relative weights among the constraints, along with the degree of change in weighting brought about by error-driven learning. The larger numbers here reflect the fact that the child may have to recognize many examples of a given error before the relevant weightings change enough to change the child's productions.

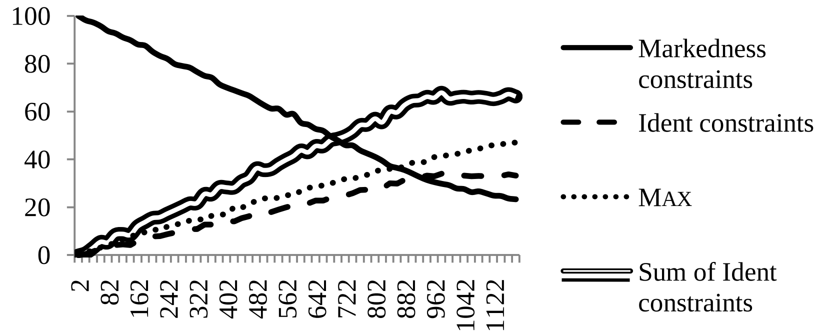

Figure 1. Trajectory of constraints learned by the HG-GLA. The single solid black line represents the average weight of the two markedness constraints, while the small dotted line represents the weight of MAX. The larger dashed line represents the average weight of the two IDENT constraints, and the double line represents their cumulative weight.

It is clear that a learner learning with the HG-GLA can acquire a CFE, but why does this happen? Free variation in the early stages of learning is crucial to the problem. Remember that outputs with equal weights are assumed to occur in free variation. At the earliest points of learning, when all of the faithfulness constraints have weights of 0, then multiple unmarked, unfaithful outputs are competing, and Amahl is producing different unmarked, unfaithful outputs. For the adult target *bed*, about half the time we would expect him to produce [b̥ɛt], and the other half of the time we would expect him to produce [b̥ɛ]. Imagine Amahl as a learner, shifting his constraint weights as he recognizes mismatches between his outputs and the adult target. When Amahl hears a word like *bed* and compares it to his own output, half the time he would increase the weight of IDENT[voice] by some small amount, and the other half of the time he would increase the weight of MAX by some small amount. The same is true for the word *bus*: when Amahl compares the adult target to his output [b̥ʌt], he would increase the weight of IDENT[continuant], but when he compares the target to his output [b̥ʌ], he would increase the weight of MAX. And upon hearing *noise*, Amahl would increase the weight of all three faithfulness constraints: IDENT[voice] and IDENT[continuant] when comparing the adult target to his output [nɔit], and MAX when comparing the adult target to his output [nɔi:].

If we add up all those increases, the reason for the CFE becomes clear. When Amahl hears *bed*, he increases the weight of IDENT[voice] and MAX. When he hears *bus*, he increases the weight of IDENT[cont] and MAX. And when he hears *noise*, he increases the weight of all three. The weight of MAX increases at 1.5 times the weight of either of the IDENT constraints. In essence, deletion is a possible repair for any of the marked structures in Amahl's grammar. That is, Amahl can delete a voiced coda, or a fricative, or a voiced coda fricative—thus when he hears any of the three relevant learning data, he must increase the weight of MAX. The same is not true of the IDENT constraints—a change in voice quality is irrelevant if the target is a voiceless fricative, just as a change in continuancy is irrelevant if the input is a voiced coda stop. The learner thus gets more evidence that the target grammar does not delete segments than evidence that features are not changed. This type of learning could not occur without the free variation in the early stages. It is while Amahl's outputs are unstable that he receives evidence about MAX and the IDENT constraints at about a 3:2 ratio, the ratio that leads to the CFE.

Thus far we have determined that the presence of a CFE depends in part on early variation. What other assumptions are crucial? Following previous claims related to the subset problem (Jesney & Tessier, 2007), we assigned the markedness constraints higher initial weights than the faithfulness constraints. It has been shown, however, that even when all constraints begin with identical initial weights, the HG-GLA will converge on a set of weightings that is appropriate for the target grammar (e.g., Boersma & Pater, 2008). However, it is important that a realistic learning model not only converge on the target, but also go through realistic learning stages on the way to convergence. The question is, then, whether the initial weighting bias used in the above simulation is necessary to achieving the CFE. In order to answer this question, the same simulation was run, but with an initial weight of 10 for all the constraints. Figure 2 shows the resulting constraint trajectories.

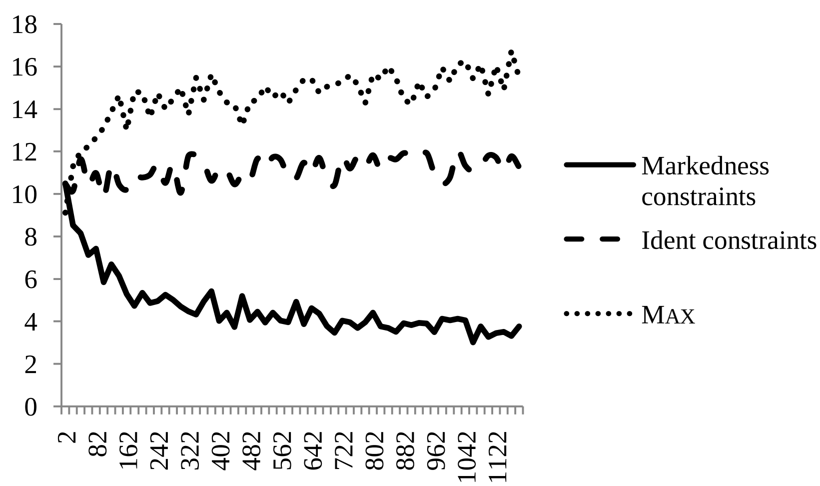

Figure 2. Trajectory of constraints when all constraints have an initial starting weight of 10. The solid black line represents the average weight of the two markedness constraints, the larger dashed line represents the average weight of the two IDENT constraints, and the small dotted line represents the weight of MAX.

When all the constraints begin with the same weight, as in Figure 2, they diverge rather quickly, and the target grammar is achieved in very few learning trials. The learner represented by Figure 2 thus converges on the target grammar, but without going through a CFE stage first. Instead, MAX (represented by the gray line) and the IDENT constraints (represented by the dotted line) quickly achieve higher weights than the markedness constraints (represented by the black line). For the CFE to occur, then, it seems that the markedness constraints must begin with a weight that is enough higher than the faithfulness constraints that the faithfulness constraints have room to diverge before overtaking the weight of the markedness constraints. In other words, the child must produce unmarked outputs long enough to go through the different unfaithful stages that are determined by the relative weighting of the faithfulness constraints.

Another assumption made by Jesney and Tessier (2007) is that, in order to solve an unrelated problem arising from morphological concatenation and learning, markedness constraints must be demoted at a faster rate than faithfulness constraints are promoted. If this assumption is true, it has implications for learning a CFE. To test this assumption, we ran the same simulation that produced a CFE, but the plasticity of the markedness constraints was set at 20 times the plasticity of the faithfulness constraints (an extreme value in order to prove the point). The result is shown in Figure 3. It is clear in that figure that the markedness constraints decline at a steep rate, such that their weights fall lower than those of the faithfulness constraints by about the 80th learning trial. When the markedness constraints decrease so rapidly, the faithfulness constraints do not have time to diverge before the target grammar is achieved. The grammar in Figure 3 thus converges on the target without having

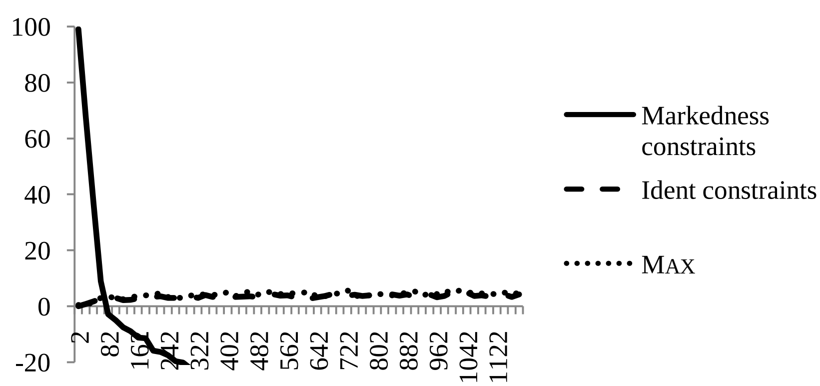

Figure 3. Trajectory of constraints when the markedness constraints have a plasticity 20 times greater than that of the faithfulness constraints. The solid black line represents the average weight of the two markedness constraints, the larger dashed line represents the average weight of the two IDENT constraints, and the small dotted line represents the weight of MAX.

gone through a CFE stage. Of course, the rate of decline of the markedness constraints in this example is extreme, but the point still holds: if the markedness constraints are assumed to decline too rapidly, the CFE stage will not emerge or will be cut short.

A final crucial assumption of the simulation shown in Figure 1 is that the learner encounters different word types with approximately equal frequency. That is, the learner hears words with voiced codas, words with fricatives, and words with voiced coda fricatives at equal rates. This is not realistic. In general, more marked sounds occur less frequently in a language sample (Greenberg, 1966; Croft, 2003; Haspelmath, 2006). We would expect, then, that Amahl would have heard words with coda [z] less frequently than words with coda [s] or [d] (though the fact that the English plural morpheme is typically realized as a coda [z] is encouraging). In order to make the simulation better match Amahl's language input, it would be important to give the model learning data that approximate the frequency of the input sounds in the speech that the child actually hears. This may change the rate at which the CFE is acquired. Even without this assumption of perfectly equal learning data, though, the fact remains that Amahl will have more opportunities to increase the weight of MAX than of either of the IDENT constraints, because of the simple fact that deletion can be used as a repair for a broader subset of the marked structures than either of the IDENT constraints (for instance, Amahl also uses segmental deletion as a repair for NC clusters). We can thus expect the weight of MAX to increase at a greater rate than that of the IDENT constraints, and, for some period of time, we can also expect that the weight of MAX will not yet overtake the sum of the IDENT constraints.

4. Summary and conclusion

We have shown that CFEs can be acquired even as a child is learning a target grammar in which no CFE exists. Such a CFE can be accounted for by appealing to weighted constraints like those in HG. Moreover, the acquisition of a CFE can be modeled in an HG-GLA learning simulation as long as several key assumptions are observed: that markedness constraints outweigh faithfulness constraints by the point at which this learning stage begins, that the weight of markedness constraints does not decline too rapidly, and that the learner hears forms with roughly equal representation of the different marked structures. The next step in this line of research is to test (and modify as necessary) these assumptions in order to determine the most realistic model of CFE acquisition.

References

Boersma, Paul. 1998. Functional Phonology: Formalizing the Interaction Between Articulatory and Perceptual Drives. The Hague, Holland Academic Graphics.

Boersma, Paul and Joe Pater. 2008. Convergence properties of a gradual learning algorithm for Harmonic Grammar. Ms., University of Amsterdam and University of Massachusetts, Amherst.

Boersma, Paul & David Weenink. 2008. Praat: Doing phonetics by computer (Version 5.0.32). [Computer program]. Retrieved August 25, 2008, from http://www.praat.org/.

Croft, William. 2003. Typology and Universals (2nd edition). Cambridge: Cambridge University Press.

Farris-Trimble, Ashley. 2008. Cumulative faithfulness effects in phonology. Unpublished doctoral dissertation, Indiana University, Bloomington.

Goldrick, Matthew and Robert Daland. In press. Linking speech errors and phonological grammars: Insights from Harmonic Grammar networks. Phonology.

Greenberg, Joseph. 1966. Language Universals, with Special Reference to Feature Hierarchies (Janua Linguarum, Series Minor 59). The Hague: Mouton.

Haspelmath, Martin. 2006. Against markedness (and what to replace it with). Journal of Linguistics 42: 25-70.

Jesney, Karen and Anne-Michelle Tessier. 2007. Re-evaluating learning biases in Harmonic Grammar. In University of Massachusetts Occasional Papers in Linguistics 36: Papers in theoretical and computation phonology, ed. by Michael Becker. Amherst, MA: GLSA Publications.

Legendre, Geraldine, Yoshiro Miyata and Paul Smolensky. 1990a. Harmonic Grammar - A formal multi-level connectionist theory of linguistic well-formedness: An application. Proceedings of the Twelfth Annual Conference of the Cognitive Science Society. Cambridge, MA, Lawrence Erlbaum: 884-891.

Legendre, Geraldine, Yoshiro Miyata and Paul Smolensky. 1990b. Harmonic Grammar - A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. Proceedings of the Twelfth Annual Conference of the Cognitive Science Society. Cambridge, MA, Lawrence Erlbaum: 388-395.

McCarthy, John J. 1999. Sympathy, cumulativity, and the Duke-of-York gambit. Optimal Green Ideas in Phonology. K. Baertsch and D. A. Dinnsen. Bloomington, IN, Indiana University Linguistics Club. Indiana University Working Papers in Linguistics 1: 57-91.

McCarthy, John J. 2002. A Thematic Guide to Optimality Theory. Cambridge, UK, Cambridge University Press.

McCarthy, John J. 2007. Hidden Generalizations: Phonological Opacity in Optimality Theory. London, Equinox.

Pater, Joe. 2008. Gradual learning and convergence. Linguistic Inquiry 39: 334-345.

Pater, Joe, Rajesh Bhatt and Christopher Potts. 2007. Linguistic optimization. Ms., University of Massachusetts, Amherst. [ROA 924.]

Pater, Joe, Karen C. Jesney and Anne-Michelle Tessier. 2007. Phonological acquisition as weighted constraint interaction. Proceedings of the 2nd Conference on Generative Approaches to Language Acquisition—North America. Ed. by Alyona Belikova, Luisa Meroni and Mari Umeda. Somerville, MA, Cascadilla Proceedings Project: 339-350.

Prince, Alan and Paul Smolensky. 1993/2004. Optimality Theory: Constraint Interaction in Generative Grammar. Malden, MA, Blackwell.

Smith, Neil V. 1973. The Acquisition of Phonology: A Case Study. Cambridge, UK, Cambridge University Press.

Smolensky, Paul and Geraldine Legendre. 2006. The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar. Cambridge, MA, MIT Press.