Modeling Doubly Marked Lags with a Split Additive Model

Adam Albright, Giorgio Magri, and Jennifer Michaels
Massachusetts Institute of Technology

1. Introduction: Marked Structures in Phonological Acquisition

In child phonology marked structures are frequently acquired first in relatively unmarked contexts. Four examples of this can be observed in the acquisition of Dutch syllable structure as reported by Levelt *et al.* (2000), a representative example of which is given in (1): onsetless syllables must initally be open (stage 3), as must syllables with complex onsets (stage 5); syllables with complex codas at first require onsets (stage 7); and complex margins require simple opposite-side margins (stage 8).

(1) A representative order of acquisition of Dutch syllable types

Stage 1	CV
Stage 2	CV, CVC
Stage 3	CV, CVC, V (***VC**)
Stage 4	CV, CVC, V, VC
Stage 5	CV, CVC, V, VC, CCV (***CCVC**)
Stage 6	CV, CVC, V, VC, CCV, CCVC
Stage 7	CV, CVC, V, VC, CCV, CCVC, CVCC (***VCC**)
Stage 8	CV, CVC, V, VC, CCV, CCVC, CVCC, VCC (***CCVCC**)
Stage 9	CV, CVC, V, VC, CCV, CCVC, CVCC, VCC, CCVCC

For example, in stage 5, children produce syllables with codas (CVC) and syllables with complex onsets (CCV), but not syllables that contain codas and complex onsets simultaneously (CCVC). A concrete example can be found in the data from Enzo in the Fikkert-Levelt CLPF corpus: from ages 1;11 through 2;4, /xrun/ 'green' is produced variably as [χun] (CVC) or [χʁu]/[χʁunə] (CCV). We refer to delays of structures such as CCVC as *doubly marked lags*, since the delayed form violates two markedness constraints, whereas the simpler syllable structures that are acquired first violate only one. Thus, in a doubly marked lag stage, marked structures are allowed individually, but cannot occur simultaneously.

(2) Delayed structures are doubly marked

CVC:	violates *Coda	(syllables should not have codas)
CCV:	violates *#CC	(syllables should not have complex onsets)
CCVC:	violates both *Coda and *#CC	

---

[*] We wish to thank Edward Flemming, Clara Levelt, Joe Pater, Donca Steriade, and audiences at MIT, the Northeast Computational Phonology Circle, and BUCLD 32 for helpful comments and suggestions. All errors and oversights are, of course, our own.

Such bans on doubly marked forms are interesting because although they appear to be common during acquisition, they are unattested in adult languages. Optimality Theory (OT; Prince & Smolensky 2004) neatly predicts the lack of such bans as follows: first, in order to allow the individually marked structures CVC and CCV, the faithfulness constraint Max (don't delete) must outrank each of the relevant markedness constraints *Coda and *#CC.

(3) Acquisition of singly marked structures

a. Acquisition of Codas

/CVC/	Max	*Coda
☞ CVC		*
CV	*!	

b. Acquisition of Complex Onsets

/CCV/	Max	*#CC
☞ CCV		*
CV	*!	

However, once Max outranks each markedness constraint individually, violations of both are immediately tolerable, and doubly marked CCVC surfaces faithfully rather than simplifying to CVC or CCV.

(4)

/CCVC/	Max	*Coda	*#CC
☞ a. CCVC		*	*
b. CV	**!		
c. CCV	*!		*
d. CVC	*!	*	

Even though CCVC is doubly marked, it wins because multiple violations of lower ranked markedness constraints cannot 'gang up' to overcome the higher ranked faithfulness constraint. Therefore, standard OT predicts the acquisition of CCVC to follow automatically from the acquisition of CVC and CCV, with no lags.

Thus, doubly marked lags pose a two-sided challenge. The first is to provide a mechanism by which lower ranked markedness constraints can gang up to ban doubly marked structures during acquisition. We propose an additive model of constraint interaction that allows markedness constraints to gang up in a way not allowed by standard OT or previous additive models. The second challenge is to explain why such interactions are seen as acquisition stages, but never in adult systems. To account for the more restricted set of constraint interactions seen in adult language, we propose a learning procedure in which additive interactions are unstable, and learners are unlikely to remain in such stages.

2. Previous accounts to penalizing doubly-marked structures

2.1. Constraint conjunction

One mechanism that allows markedness constraints to gang up and outrank faithfulness, even when neither does alone, is constraint conjunction (Kirchner 1996; Levelt *et al.* 2000). This approach maintains the standard OT model of constraint interaction, but enriches the constraint set with conjoined constraints such as *CODA∧*#CC. This constraint is violated only by structures with both codas and complex onsets in the same syllable—that is, only by CCVC. By ranking *CODA∧*#CC above MAX, but MAX above *CODA and *#CC individually, we correctly capture the doubly marked lag stage in which CVC and CCV are both possible, but CCVC is not.

(5) Ranking: (*CODA∧*#CC) ≫ MAX ≫*CODA, *#CC

/CCVC/	*CODA∧*#CC	MAX	*CODA	*#CC
CCVC	*!		*	*
CV		**!		
☞ CCV		*		*
☞ CVC		*	*	

More generally, it is possible to construct conjoined constraints that specifically ban each of the doubly marked structures for which lags are observed in (1).

(6) Conjoined constraints banning doubly marked structures

VC	*CODA∧ONSET	CCVC	*CODA∧*#CC
VCC	ONSET∧*CC#	CCVCC	*#CC∧*CC#

However, positing separate constraints for each doubly marked structure also poses a learning issue: it is not clear that learners must overcome all of these pairwise restrictions independently. As we will show in section 4.3, improved performance on simple syllables with codas (CVC) appears to go hand in hand with improved performance on more marked syllables involving codas. In other words, we need a model in which doubly marked structures are ruled out by the same constraints that ban their subparts.

2.2. Additive evaluation

A different approach to modeling ganging up effects is to stick with the standard set of syllable structure constraints, but to allow the constraints to interact additively (Legendre *et al.* 1990; Keller 2000; Goldwater & Johnson 2003; Pater *et al.* 2007b; Pater *et al.* 2007a; Jäger, in press). In an additive (or linear) model, constraints are assigned numeric weights rather than strict rankings. The *penalty* of a candidate is determined by summing its weighted violations, and the candidate with the smallest penalty wins ((7a)). Since the penalty is a sum over all constraints, candidates with multiple violations of lower-weighted constraints may receive a greater penalty than candidates with fewer violations of higher-weighted constraints. This ganging up of lower weighted constraints is illustrated schematically in ((7b)).

(7) Evaluating candidates by summing weighted violations

a. Violations of higher-weighted constraints lead to greater penalties

/UR/	C_1	C_2	C_3	Penalty
Weight:	3	2	1.5	
candidate$_1$	1			3
candidate$_2$		1		2
☞ candidate$_3$			1	1.5
candidate$_4$			2	3

b. Ganging up of lower-weighted constraints

/UR/	C_1	C_2	C_3	Penalty
Weight:	3	2	1.5	
☞ candidate$_1$	1			3
candidate$_2$		1	1	3.5

The additive approach to ganging up effects is appealing because it makes use of two independently needed constraints, rather than relying on an additional conjoined constraint. Unfortunately, additive evaluation by itself is not sufficient to capture the type of ganging up effect seen in doubly marked lags, for reasons discussed by Pater *et al.* (2007a). As (8) shows, in order for CCV to surface faithfully, the weight of MAX (designated here as f) must be greater than the weight of *#CC (designated m_1). Similarly, in order for CVC to surface faithfully in (9), the weight of MAX must be greater than the weight of *CODA (designated m_2). As seen in (10), these conditions entail that CCVC must also surface faithfully.

(8) CCV surfaces faithfully if $f > m_1$

/CCV/	MAX	*#CC	*CODA	Penalty
Weight:	f	m_1	m_2	
☞ a. CCV		*		m_1
b. CV	*			f

(9) CVC surfaces faithfully if $f > m_2$

/CVC/	MAX	*#CC	*CODA	Penalty
Weight:	f	m_1	m_2	
☞ a. CVC			*	m_2
b. CV	*			f

(10) CCVC must surface faithfully if $f > m_1$ and $f > m_2$

/CCVC/	MAX	*#CC	*CODA	Penalty
Weight:	f	m_1	m_2	
☞ a. CCVC		*	*	$m_1 + m_2$
b. CVC	*		*	$f + m_2$
c. CCV	*	*		$f + m_1$
d. CV	**			$f + f$

Candidate (10b) loses to (10a), since $f > m_1$ (from (8)); likewise candidate (10c) loses, since $f > m_2$ (from (9)). Candidate (10d) is worst of all, incurring two faithfulness violations. Therefore, the conditions in (8–9) are incompatible with banning the doubly marked structure, and preclude a straightforward analysis of doubly marked lags as an additive effect. (For discussion, see Pater *et al.* 2007a.)

3. A new approach to doubly marked lags in child language

Recall that in the CCVC lag stage, codas and complex onsets are independently permitted, but the doubly marked CCVC is simplified to CVC or CCV. Comparing the candidates in table (10), we see that the unattested candidates CCVC (10a) and CV (10d) involve either two markedness violations ((10a)) or two faithfulness violations ((10d)). The preferred candidates CVC (10b) and CCV (10c) both incur a mix of markedness and faithfulness violations. The intuition that we pursue here is that multiple violations of the same constraint type exacerbate one another in a way that violations of different constraint types do not.

3.1. The Split Additive Model (SAM)

In order to capture this intuition, we propose a *split additive model* (SAM). As in the simple additive model, candidates are assigned weighted sum of violations. However, in the split model, weighted faithfulness and markedness penalties are tallied separately, and candidates are assigned a total penalty equal to the greater of the two sums. As above, the winning candidate is the one with smallest total penalty.

(11)
$$\text{Penalty(output)} = \text{argmax} \left\{ \begin{array}{l} \text{weighted sum of } \mathscr{F} \text{ violations,} \\ \text{weighted sum of } \mathscr{M} \text{ violations} \end{array} \right\}$$

Given a suitable constraint weighting, the split additive model can capture doubly marked lags. When the weight of MAX is less than the sum of the weights of two markedness constraints, a candidate with multiple markedness violations (CCVC) will lose to a candidate which distributes its violations across markedness and faithfulness (CVC or CCV).

(12) Separate summation of faithfulness and markedness

/CCVC/	MAX $f=3$	*#CC $m_1=2$	*CODA $m_2=2$	$\mathscr{F}$ Penalty	$\mathscr{M}$ Penalty	Total
CCVC		*	*	0	**4**	**4**
☞ CVC	*		*	**3**	2	**3**
☞ CCV	*	*		**3**	2	**3**
CV	**			**6**	0	**6**

In short, separate summation of markedness and faithfulness violations allows markedness constraints to gang up in a way that OT and the simple additive model do not. This allows us to capture acquisition stages like the ones that Levelt et al. observe, given the right set of weights. What remains to be shown, however, is whether there is some learning procedure that would actually pass through such stages on the way to adult Dutch. We now demonstrate such a procedure by running a standard error-driven learning algorithm which implements SAM.

3.2. Testing the model

We implemented an error-driven learning model, which has the following steps. First, the model "hears" an adult surface form, sampled from the target language. The attested adult form (the *target* form) is then compared against the surface form favored by SAM using the current set of constraint weights (the *predicted* form). If the current grammar correctly predicts the adult form (*predicted* = *target*), the model moves on to the next datum. If the current grammar does not predict that adult form (*predicted* $\neq$ *target*), then the weights of all constraints are adjusted according to the rule in (13).

(13) Update rule for the weight of constraint *c*

1. Calculate difference in violations between predicted and target forms
   $\textit{Violation difference} = \textit{violations}_{\mathrm{c}}\textit{(predicted)} - \textit{violations}_{\mathrm{c}}\textit{(target)}$
2. Adjust the weight of constraint *c* by the difference[1]
   $\textit{New weight} = \textit{old weight} + \textit{violation difference}$

The rule in (13) has the following effect: if a constraint assigns more violations to the predicted form than to the target form, the weight of the constraint will be incremented in hopes of ruling out the incorrectly predicted candidate. Conversely, if a constraint assigns more violations to the target than to the predicted form, its weight is decremented so that the target form is penalized less.

We assume the set of syllable structure constraints in (14).

(14) a. Markedness:

ONSET	Syllables must have onsets
*CODA	Syllables may not have codas
*#CC	Syllables may not have complex onsets
*CC#	Syllables may not have complex codas

b. Faithfulness:

DEP	Don't insert segments
MAX(C)/___ [+son]	Preserve consonants before sonorants
MAX(C)/V___	Preserve consonants after vowels
MAX(C)/C___ #	Preserve consonants in final clusters

---

[1] Technically, weights are adjusted by an amount equal to the violation difference multiplied by a scaling factor η, the *plasticity*. In the simulations reported here, the plasticity is set to 1 and not altered during the course of training.

We further assume that the initial weights of constraints reflect two biases. The first bias is that markedness constraints start out weighted higher than faithfulness constraints (Smolensky 1996). The second bias is that faithfulness constraints that refer to perceptually privileged positions are weighted higher than those that refer to perceptually weaker positions (Steriade 2001). Specifically, we adopt the three-tiered initial weighting scheme in (15). The precise values are somewhat arbitrary; what is crucial is the relative position of markedness with respect to different types of faithfulness.

(15) Initial weights assumed

Markedness:	100
"Privileged" Faithfulness:	30
"Unprivileged" Faithfulness:	15

We trained the model on an input of Dutch syllable structures, sampled according to their frequencies in child-directed Dutch (Levelt et al 2000).

(16) Frequencies of syllable types in input data

CV	44.90%	V	3.85%	CCV	1.38%
CVC	33.05%	CVCC	3.25%	VCC	.42%
VC	11.99%	CCVC	1.98%	CCVCC	.26%

Under these learning conditions, the model converges to adult Dutch (i.e., all adult forms are produced correctly) after approximately 2100 trials. Furthermore, when we examine the model's time course of acquisition, we find that syllable types emerge in the following order.

(17) CV > CVC > V > VC > CVCC > CCV > CCVC > VCC > CCVCC

We observe that the model correctly predicts all four doubly marked lags that are seen in the Levelt et al. data (V > VC; CVCC > VCC; CCV > CCVC; CCV, CVCC > CCVCC). This confirms that an implemented learning model using SAM does indeed pass through stages in which additive interactions lead to doubly marked lags. Note that there are also discrepancies between the order here and the empirically observed order in (1) above; we return to these in section 4.3.

4. Lack of doubly-marked bans in adult grammars

As we've seen above, without a device such as constraint conjunction, OT and simple additive models cannot produce grammars in which singly marked structures are allowed but doubly marked structures are banned. In fact, this is a design feature, intended to account for the typological fact that in adult languages, the existence of singly marked structures appears to entail the existence of doubly marked structures.[2] This produces a quandary: in order to account for doubly marked lags in children, we have adopted a more powerful system which allows markedness constraints to gang up additively to overcome faithfulness. However, this predicts that adult systems, too, should be able to ban doubly marked structures. The challenge, then, is to explain why such systems are unattested.

In an additive model, singly marked structures emerge when the weights of markedness constraints pass below weights of faithfulness constraints—e.g., $w(\textsc{Max}) > w(\textsc{*Coda})$. Crucially, doubly marked lags require a very special configuration of weights: on the one hand, the weights of the two markedness constraints must be low enough to be less than faithfulness, but on the other hand, they must be high enough that their sum outweighs faithfulness.

(18) $$w(\textsc{*Coda}) + w(\text{*\#CC}) > w(\textsc{Faith}) > \left\{ \begin{array}{c} w(\textsc{*Coda}) \\ w(\text{*\#CC}) \end{array} \right\}$$

The basic learning trajectory is for the weights of markedness constraints to decrease, and the weights of faithfulness constraints to increase. Observe that if $w(\textsc{*Coda})$ and $w(\text{*\#CC})$ continue to decrease, their sum will fall below $W(\textsc{Faith})$, and additive interaction will no longer be possible. We wish to derive this effect by modifying the learning algorithm so that learning continues until additive interactions no longer occur.

4.1. Aggressive error-driven learning

Recall that in the simple error-driven model described in section 3.1, we demand, given a datum CVC, that the grammar prefer a faithful mapping of /CVC/ to [CVC]. Using this procedure, as soon as $w(\textsc{Max})$ exceeds $w(\textsc{*Coda})$ by even a tiny amount, the difference is sufficient to yield a preference for CVC, and further exposure to CVC ceases to cause further reweighting. It is possible to adopt a more ambitious form of error-driven learning, however, which demands that the grammar not only prefer CVC, but guarantees CVC at a rate approaching 100% (to some tolerance threshold). In order to guarantee a categorical preference, the weights must end up farther apart than simple error-driven learning would require. As a consequence, once the difference between the weights of faithfulness and markedness becomes quite large, it will be more difficult for markedness constraints to gang up and ban doubly marked structures. Thus, it appears that a more aggressive form of error-driven learning may eventually stamp out doubly-marked lags.

---

[2] Levelt & van de Vijver (2004) mention one possible example of a language with V and CVC but not VC: Central Sentani (Hartzler 1976). In fact, Hartzler lists several examples of VC (*em fæu* 'banana', *miæ aŋ* 'bachelor', *riaʔ* 'meeting', but comments that they are rare. We believe that the rarity of VC in Central Sentani may reflect a conspiracy of independent constraints (codas occur only word-finally, medial hiatus is avoided). However, further investigation is needed to determine whether Sentani—or, indeed, any adult language—shows a significant gradient dispreference for doubly marked syllable types.

We implemented this idea as follows: on receiving input data, we check that model produces the correct target output with a probability of essentially 100%.[3] The probability of an output form, given the current weights and a set of competing candidates, is defined as in (19), as is standard in maximum entropy modeling (Goldwater & Johnson 2003; Hayes and Wilson, in press).

(19) $$\text{Prob(output)} = \frac{e^{-\text{penalty(output)/T}}}{\sum\limits_{\text{candidates}} e^{-\text{penalty(candidate)/T}}}$$

- Total penalties calculated using SAM (described above)
- T = a scaling factor (here, 2)

If the probability of the given adult target form is less than 100%, we adjust the constraint weights as if the model has made an error.[4]

4.2. Testing aggressive error-driven learning

We re-ran the learning simulation, using the more aggressive form of error-driven learning described in the previous section. In Figure 1, we show the initial learning stages through the acquisition of the first doubly marked lag (VC). We see that *CODA is demoted quickly, since almost half of the input tokens involve codas (CVC, VC, CCVC, CVCC, VCC, CCVCC). The next most frequent set of inputs are vowel-initial forms (V, VC, VCC), causing ONSET to be demoted relatively quickly as well. This leads to a stage (starting around trial 180) in which CVC and V are both possible, but the combined weights of ONSET and *CODA are still greater than the weights of either MAX or DEP. This is a doubly marked lag stage, in which VC is realized unfaithfully as CVC. However, the weight of ONSET continues to fall, and the lag state ends by around trial 360.

The emerge of VC in this simulation is much faster than would be produced by training on VC alone; it is continued exposure to singly marked CVC forms that motivates the continued rapid decrease in the weight of *CODA. This is the desired result: training on simple forms helps to eventually eliminate ban on doubly marked forms. Furthermore, this result holds even if doubly marked structures are completely absent from the input. We ran a subsequent simulation in which VC forms were completely withheld from the learner, and the results were virtually identical to those in Figure 1. This makes the correct typological prediction—namely, that doubly marked lags do not persist past the early learning stages.

The analysis presented here takes a rather different approach to the relation between grammar and typology than is typical in the OT literature. In our account,

---

[3] Specifically, to a probability within one hundred trillionth of a percent (which is the precision limit of our Perl implementation).

[4] Jäger (in press) adopts a similar approach, in which the predictions of the current set of weights are sampled probabilistically, occasionally generating an error if the probability of the target form is less than 100%.

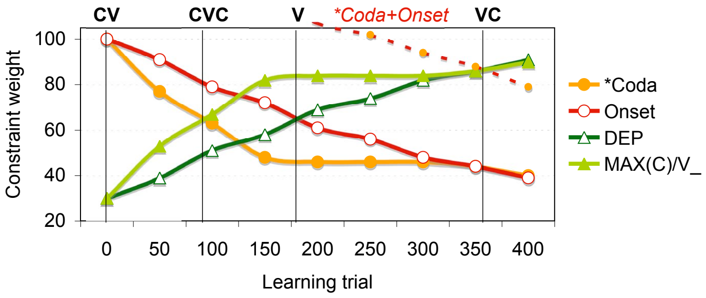

Figure 1: Learning trajectory, using aggressive error-driven learning

doubly marked lag states are grammatically possible languages, but learners will not converge on them without positive evidence that the doubly marked structures exist underlyingly, but are eliminated. Thus, the absence of such languages as adult systems is attributed not to the architecture of the grammar itself, but to the nature of the learning mechanism.

4.3. Hot on the heels order of acquisition

Another interesting aspect of the acquisition order reported by Levelt et al. (2000) is that although it is motivated in part by frequency, there are also instances in which very infrequent structures are acquired before structures that are somewhat more frequent. For example, children who acquire complex onsets before complex codas (the upper path in (20)) produce CCVC before CVCC, even though it is significantly less frequent (1.98% vs. 3.25%). In other words, after mastering complex onsets in simple contexts (CCV), children next move on to acquire complex onsets in more marked contexts (CCVC). We call this *hot on the heels* order of acquisition.

(20) Hot on the heels acquisition of CCVC and VCC

$$\mathrm{CV} > \mathrm{CVC} > \mathrm{V} > \mathrm{VC} > \left\{ \begin{array}{c} \boxed{\mathrm{CCV} > \mathrm{CCVC}} > \boxed{\mathrm{CVCC} > \mathrm{VCC}} \\ \boxed{\mathrm{CVCC} > \mathrm{VCC}} > \boxed{\mathrm{CCV} > \mathrm{CCVC}} \end{array} \right\} > \mathrm{CCVCC}$$

As shown in (17), a simple error-driven algorithm (or, indeed, any frequency-based learning model) predicts that CVCC will be acquired early, while VCC will be acquired late, owing to its low frequency. The challenge is to explain why VCC is acquired hot on the heels of CVCC.

It turns out that the aggressive error-driven approach described above helps to model hot on the heels acquisition. In this model, further training on simple structures continues to drive learning until success is guaranteed. For example, further exposure to CVC, CCV continue to cause demotion of *CODA, *#CC.

This yields the prediction that the order of acquisition of doubly marked structures depends on the frequency of their simple counterparts, rather than solely on the surface frequency of the doubly marked structures themselves. Concretely, if CVC, CCV are frequent enough to demote *Coda, *#CC quickly, then they will continue to motivate demotion and CCVC will follow hot on their heels.

In order to test the usefulness of this mechanism in predicting hot on the heels behavior, we ran 20 simulated childhoods, training each according to the child-directed input frequencies above. In 15/20 runs, marked structures correctly emerged in pairs, resolving the incorrect ordering of CVCC with respect to CCV seen in (17).

(21) Hot on the heels order

a. {CCV, CCVC} > {CVCC, VCC} (14 times)[5]

b. {CVCC, VCC} > {CCV, CCVC} (1 time)

In the remaining 5 trials, CVCC emerged "early" (i.e., with CCV and CCVC intervening between CVCC and VCC), reflecting the higher frequency of CVCC. Our provisional conclusion is that although the mechanism of aggressive error-driven learning does not eliminate frequency-based order of acquisition altogether, it does succeed in predicting hot on the heels order of acquisition much of the time.

5. Summary and conclusion

In this paper, we have proposed a split additive model which allows ganging up effects, but only among constraints of the same type (markedness, faithfulness). We have shown that this model can capture doubly marked lag stages. We attributed the absence of such stages from adult systems to the dynamics of the learning model. Specifically, since doubly-marked structures are penalized by the same constraints as singly-marked structures, continued exposure to simpler forms eventually drives the weights of the relevant constraints too far apart to allow additive interactions. We have suggested that this also helps to explain the relative rapidity of acquisition of doubly marked structures.

The model presented here leaves a number of open issues. First, markedness penalties involve summation of all markedness violations. In principle, this predicts a wide array of markedness interactions within the word. Here, we have investigated interactions of just one kind of constraint (syllable structure) within a local domain (a syllable). It is entirely possible that additive interactions are confined to particular subsets of constraints, or violations within a restricted domain.

---

[5] The fact that CCV emerges first in the majority of simulations is due to statistics of this particular child-directed corpus. Levelt et al. (2000) actually observe a bias towards the opposite order ((21b)). However, the frequency differences between complex onsets and complex codas are very small, and a small sampling difference could reverse the preference. We therefore believe that the discrepancy may be due to the fact that this specific corpus is not representative of the experience of every Dutch-learning child.

At the same time, we are struck by the fact that child phonology often involves simplification in contexts that are marked in a possibly "irrelevant" way—e.g., segmental simplifications in marked prosodic contexts (Pater *et al.* 2007b), or in more complex morphological contexts (e.g., Dinnsen & McGarrity 2004). Further empirical data needed to determine extent and limits of additive interactions.

A second issue is that separate summation of markedness and faithfulness violations predicts the possibility of lags in doubly unfaithful structures. In fact, these do occur, in the form of chain shifts: for example, in a stage where *puddle* is produced as *puggle* ($*dl > \mathscr{F}$(place)), *puzzle* may be produced as *puddle* (violating $\mathscr{F}$(strident)), rather than doubly-unfaithful *puggle* ($\mathscr{F}$(strident) $+$ $\mathscr{F}$(place)) (Smith 1973). The application of SAM to chain shifts in acquisition is left as a matter for future research.

References

Dinnsen, Daniel, & Laura W. McGarrity. 2004. On the nature of alternations in phonological acquisition. *Studies in Phonetics, Phonology and Morphology* 10.23–41.

Goldwater, Sharon, & Mark Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. In *Proceedings of the Workshop on Variation within Optimality Theory, Stockholm University*, ed. by Jennifer Spenader, Anders Eriksson, & Osten Dahl, 111–120. Stockholm University.

Hartzler, Margaret. 1976. Central Sentani phonology. *Irian* 5.66–81.

Hayes, Bruce, & Colin Wilson. In press. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* .

Jäger, Gerhard. In press. Maximum entropy models and Stochastic Optimality Theory. In *Architectures, Rules, and Preferences: A Festschrift for Joan Bresnan*, ed. by Jane Grimshaw, Joan Maling, Chris Manning, Jane Simpson, & Annie Zaenen. Stanford: CSLI Publications.

Keller, Frank. 2000. *Gradience in grammar: Experimental and computational aspects of degrees of gramaticality*. Univ. of Edinburgh dissertation.

Kirchner, Robert. 1996. Synchronic chain shifts in Optimality Theory. *Linguistic Inquiry* 27.241–350.

Legendre, Geraldine, Yoshiro Miyata, & Paul Smolensky. 1990. Harmonic grammar: A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. Technical Report 90-5, Institute of Cognitive Science, Univ. of Colorado.

Levelt, Clara, Niels O. Schiller, & Willem J.M. Levelt. 2000. The acquisition of syllable types. *Language Acquisition* 8.237–264.

——, & Ruben van de Vijver. 2004. Syllable types in cross-linguistic and developmental grammars. In *Constraints in Phonological Acquisition*, ed. by René Kager, Joe Pater, & Wim Zonneveld, 204–218. Cambridge: Cambridge University Press.

Pater, Joe, Rajesh Bhatt, & Christopher Potts, 2007a. Linguistic optimization. University of Massachusetts, Amherst ms.

——, Karen Jesney, & Anne-Michelle Tessier. 2007b. Phonological acquisition as weighted constraint interaction. In *Proceedings of the Conference on Generative Approaches to Language Acquisition North America (GALANA 2)*, ed. by Alyona Beliokova, Luisa Meroni, & Mari Umeda, Somerville, MA. Cascadilla Proceedings Project.

Prince, Alan, & Paul Smolensky. 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Blackwell Publishing.

Smith, Neilson V. 1973. *The Acquisition of Phonology*. Cambridge University Press.

Smolensky, Paul. 1996. The initial state and 'Richness of the Base' in Optimality Theory. Technical report, Johns Hopkins University.

Steriade, Donca. 2001. Directional asymmetries in place assimilation: a perceptual account. In *Perception in Phonology*, ed. by Elizabeth Hume & Keith Johnson, 219–250. Academic Press.