Workshop: Acquisition of Phonology
Radboud University, Nijmegen
May 13, 2007

Gradual Learning and Gradient Phonotactics

Joe Pater, University of Massachusetts, Amherst

Gradual learning as gradual adjustment of constraint values (Rosenblatt 1962, Boersma 1998) automatically yields weighting values that can be used in an account of gradient phonotactics, if the grammar is formulated in terms of Harmonic Grammar (HG; Legendre, Smolensky and Miyata 1990). I propose a new metric for gradient acceptability in HG, and show how a complex case of gradient phonotactics, the Muna restriction against sequences of homorganic consonants (Coetzee and Pater 2007) can be analyzed in these terms.

1. Gradual learning and the GLA

The Gradual Learning Algorithm (GLA; Boersma 1998, Boersma and Levelt 2000, Boersma and Hayes 2001) is a gradual error-driven learning algorithm originally developed for Optimality Theory (OT; Prince and Smolensky 1993/2004), and extended to HG in Boersma and Weenink (2006) and Pater, Jesney and Tessier (2007)

When an error-driven learner like the GLA is supplied with an input-output pair the following steps are applied:

(1) *Error-driven learning*
    1. Parsing:
        The input is submitted to the current grammar to generate an output
    2. Comparison:
        If the resulting output differs from the learning datum, learning is triggered
    3. Learning:
        The grammar is adjusted so as to prefer the correct form (the Winner) over the learner's output (the Loser)

The GLA assumes a theory of grammar in which the relative strength of constraints is represented in terms of numerical values, or weights. It was originally developed for a stochastic version of OT (Boersma 1998) in which numerical ranking values are converted to an OT ranking each time the grammar is used for parsing. It adjusts ranking values as in (2).

(2) Subtract n from the ranking value of all constraints violated by the Winner, add n to all constraints violated by the Loser

It very similar to the Perceptron model (Rosenblatt 1962), a general purpose online error-driven learner whose update rule could be translated into the present context as in (3):

(3) Add n*(x-y) to the value of every constraint
Where $0 < n > 1$, x = Loser’s violation marks and y = Winner’s violation marks

One nice property of gradual learning in the linguistic context is that it can model the gradual process of language acquisition, in particular frequency-based orders of acquisition. Boersma and Levelt (2000) make use of this aspect of the GLA in modeling Levelt and van de Vijver's (2004) data on the acquisition of Dutch syllable structure. Here I replicate a portion of their result using the Praat (Boersma and Weenink 2006) implementation of the GLA, with Levelt and van de Vijver’s settings (n=0.1 in (2)).

Levelt, Schiller and Levelt (2000) and Levelt and van de Vijver (2004) observe that all 12 children in the Fikkert/Levelt corpus acquire codas before consonant clusters:

(4) Stage 1: CV only
Stage 2: CV and CVC
Stage 3: CV, CVC, CCVC, CVCC

They further observe that this order of acquisition correlates with the frequency of syllable types in a van de Weijer's (1999) corpus of child-directed speech containing 112,926 primary stressed syllables:

(5) *Percentage of syllables with coda and onset clusters in Dutch child-directed speech*

Codas	49.95 %
Onset clusters	3.62 %

Boersma and Levelt (2000) assume an initial state with markedness constraints at a value of 100 and the faithfulness constraints at a value of 50, corresponding to the OT initial ranking (Smolensky 1996; Gnanadesikan 2004). The relevant constraints in the initial ranking:

(6)

*COMPLEX	100
NOCODA	100
MAX	50

*COMPLEX	'No consonant clusters'	*CCV
NOCODA	'Syllables end in a vowel'	*CVC
	'Syllables begin with a consonant'	*V
MAX	'Every element of the input must be present in the output' (penalizes deletion)	*/CVC/ → [CV]

A learning datum with a cluster and a coda, *praat* 'speak':

(7) /pra:t/ → [pra:t]

Submitted to initial state grammar (*COMPLEX, NOCODA >> MAX):

(8) Learner's parse at initial state

/pra:t/	NOCODA	*COMPLEX	MAX
☞ [pa:]			* *
[pra:t]	*	*	
[pa:t]	*		*
[pra:]		*	*

Winner/Loser pair

/pra:t/ 'brush'	NOCODA	*COMPLEX	MAX
L [pa:]			* *
W [pra:t]	*	*	

The rankings values are updated by subtracting 0.1 from NOCODA and *COMPLEX and adding 0.1 to MAX, with the resulting grammar in (9).

(9)

*COMPLEX	99.9	/pra:t/ → [pa:]
NOCODA	99.9	
MAX	50.1	

The learner will continue to make parsing errors until sufficient data accumulate to push the Markedness and Faithfulness constraints into the reverse order.

The learner is therefore gradual; it is also sensitive to frequency, since the speed at which a Markedness constraint's value changes depends on how often violations are encountered in learning trials. After 800 trials with data distributed as in (5), the grammar is at the intermediate stage where it parses codas but not clusters faithfully:

(10)	*COMPLEX	95.89	/pra:t/ → [pa:t]
	MAX	84.92	
	NOCODA	75.21	

After 2400 trials, it parses both structures faithfully, as in the adult language:

(11)	MAX	94.43	/pra:t/ → [pra:t]
	*COMPLEX	92.83	
	NOCODA	75.21	

The learner will stop making errors when the faithfulness constraint is weighted (sufficiently) above the markedness constraints.

The stochastic OT account of variation:

(12) Each time the grammar is used, the ranking value is perturbed by noise (chosen from a normal distribution around its value).

Because stochastic OT converts ranking values to an OT grammar, it does not yield an account for another type of gradience - gradient phonotactics (Frisch *et al*. 2004):

(13) If two structures are parsed faithfully, with no variation, they are treated as equivalent by the grammar, regardless of their relative frequency in the words of the language

For example, an English nonce word that begins with [pw] (e.g. *Puerto Rico*) would be treated as equivalent to one with [pl] (as would ones with [mw] and [ml]). However, a gradual learner like the GLA does yield an account of gradient phonotactics if it is implemented in HG, rather than OT.

2. Gradient well-formedness in Harmonic Grammar

Harmonic Grammar (HG; Legendre Legendre, Miyata and Smolensky 1990, 2006; Smolensky and Legendre 2006) is similar in many ways to its more famous sibling Optimality Theory (OT Prince and Smolensky (1993/2004):

(14) i. Candidate representations are evaluated by constraints
ii. Where constraints conflict, the outcome is determined by their relative strength

The difference:

(15) In HG the constraints are on a numerical scale

I assume a version of weighted evaluation that Prince and Smolensky (1993/2004) suggest as a strawman; see also Flemming (2001), Prince (2002), Keller (2006), and Pater, Potts and Bhatt (2006):

(16) i. Violations are multiplied by weights
ii. A candidates' harmony (H) is the sum of the weighted violations
iii. The optimal candidate has the lowest H (if weights are positive)

Here is the initial state tableau from (8) in HG terms:

(17) Learner's parse at initial state

	100	100	50	H
/pra:t/	NoCoda	*Complex	Max	
☞ [pa:]			* *	100
[pra:t]	*	*		200
[pa:t]	*		*	150

As has been noted in earlier work (e.g. Legendre *et al.* 1990, 2006, Keller 2006, Hayes and Wilson to appear), degrees of well-formedness can be expressed in terms of Harmony scores.

(18) *Gradient well-formedness in Harmonic Grammar*

Weight	100	90	10	H
/pwej/	Dep	OCP-Lab-Ons	*Complex	
☞ [pwej]		*		90
[pəwej]	*			100

Weight	100	90	10	H
/pluw/	Dep	OCP-Lab-Ons	*Complex	
☞ [pluw]			*	10
[pəluw]	*			100

Even though both [pweij] and [pluw] are parsed faithfully, [pwej] has a lower H-value, and is therefore more harmonic than [pluw].

However, simply mapping H-values to a well-formedness scale leads to problems in an optimization system with conflicting constraints, since in many cases an ungrammatical form will have a score equal to or lower than a grammatical one (see Boersma 2004 for a syntactic example). Here's an example using Lombardi's (1999) analysis of coda devoicing:

(19) *A problem for direct mapping of H-values to well-formedness*

Weight	3	2	1	H
/pad/	Ident-Ons-Voice	*Voice	*Ident-Voice	
☞ [pat]			*	1
[pad]		*		2

Weight	3	2	1	H
/bat/	Ident-Ons-Voice	*Voice	*Ident-Voice	
☞ [bat]		*		2
[pat]	*			3

The candidate [pad] is ungrammatical, but receives the same score as grammatical [bat] (note that they would also be equivalently ordered in Coetzee's 2004 OT rank ordering approach).

Like the decision of whether an output is grammatical, the assignment of well-formedness scores must be made relative to the input:

(20) Acceptability(x) = H(y) - H(x)
Where *y* is the most harmonic alternative candidate for the same input as *x*

The result of applying this metric is that ungrammatical forms have negative scores, and grammatical forms have positive scores, as illustrated in (21)

(21) **Acceptability([pad])** = H([pat]) - H([pad]) = 1 - 2 = **-1**
Acceptability([bat]) = H([pat]) - H([bat]) = 3 - 2 = **1**

Gradient well-formedness using this metric:

(22) **Acceptability([pwej)** = H([pəwej]) - H([pwej]) = 100 - 90 = **10**
Acceptability([pluw]) = H([pəluw]) - H([pluw]) = 100 - 10 = **90**

In the next section, I show how gradual learning and this approach to gradient acceptability can be combined to yield an account of a complex case of gradient phonotactics.

3. Muna gradient phonotactics in Harmonic Grammar with gradual learning

Muna is a Western Austronesian language spoken off the coast of Sulawesi described by van den Berg (1989). In his grammar, he examines the co-occurrence restrictions on the consonants in a set of 1100 CVCV roots. He notes an absolute ban on pairs of prenasalized consonants, as well as restrictions on homorganic consonants. To further examine the OCP-Place restrictions, Coetzee and Pater (2007) compiled a list of all of the sequences of two consonants from a set of 5884 (V)CVCV(CV) roots in an electronic version of van den Berg and Sidu's (1996) Muna dictionary. This yielded a total of 7892 pairs.

We used the standard O/E measure (e.g. Frisch *et al.* 2004) to determine if a sequence occurs less often than would be expected if they occurred at random:

(23) Observed = the observed number of a sequence (e.g. t d = 3)

Expected = the number expected if they combined at random (e.g. t d = 8.97)

O/E = ratio of observed to expected (e.g. t d = 0.33); values lower than 1 indicate *under-representation*

The following table sums the observed and expected values for all of the sequences that agree and differ in terms of their place of representation. We omit identical consonants from this and following tables, since they co-occur freely, as well as pairs of prenasals, which fail to occur in a single item in our word list.

	Labial			**Coronal**			**Dorsal**		
	b p ɓ f ^{m}p ^{m}b m w			t d ɗ s ^{n}t ^{n}d ^{n}s n l r			k g ŋ ^{ŋ}k ^{ŋ}g ʁ		
	O	E	**O/E**	O	E	**O/E**	O	E	**O/E**
Labial	132	441.9	**0.30**						
Coronal	2741	2154.37	**1.27**	1338	1686.0	**0.79**			
Dorsal	875	770.9	**1.14**	1780	1527.6	**1.17**	29	163.6	**0.18**

Table 1: Adjacent non-identical homorganic and heterorganic consonants in Muna

Homorganic pairs are underrepresented, and coronals are less underrepresented than the other places (see Coetzee and Pater 2007 for stats).

In place co-occurence in the Arabic verbal roots, it has been noted that the degree of representation of coronals depends on whether they also agree in sonorancy, continuancy, and voicing (Greenberg 1950, McCarthy 1988, 1994, Pierrehumbert 1993, Frisch, Pierrehumbert and Broe 2004). Pairs of coronals that agree for these subsidiary features are less well represented than those that differ.

Similarity-based gradience extends across all three places of articulation in Muna. This can be seen in the following chart, which compares the O/E values for homorganic fricatives, stops and nasals at each place of articulation.

	Coronal	**Labial**	**Dorsal**
Voiced Stop + Voiceless Stop	d-t: 0.60 datu 'cross-eyed'	b-p: 0.10 pabu 'ineffective'	g-k: 0.07 kagala 'anklet'
Nasal + Voiced Stop	n-d: 0.25 da:no 'true'	m-b: 0.07 bomu 'bomb'	ŋ-g:0.00 –
Nasal + Voiceless Stop	n-t: 0.70 tuna 'twig'	m-p: 0.39 mopi 'loser in game of Jacks'	ŋ-k: 0.10 kaŋia 'starfish'
Nasal + Fricative	n-s: 1.17 nasa 'tired'	m-f: 1.04 mafaka 'agreement'	ŋ-ʁ: 0.00 –
Fricative + Voiced Stop	s-d: 0.55 sida 'happen'	f-b: 0.58 febuni 'hide oneself'	ʁ-g: 0.00 –
Fricative + Voiceless Stop	s-t: 0.37 tisore 'run aground'	f-p: 0.07 fopanto 'throw down'	ʁ-k: 0.40 kaʁa 'choking sound'

Table 2 O/E-values for Muna voiced and voiceless stops, fricatives, and nasals

Particularly striking in this table is the influence of [voice] agreement:

(24) a. Nasal-voiced stop pairs are more restricted than either nasal voiceless stop or stop-stop pairs (which also disagree in voice)
b. Stop-fricative pairs are more restricted when they agree in [voice] than when they don't (note that [ʁ] is a *voiced* uvular fricative).

Also evidence for the role of continuancy agreement:

(25) a. The nasal-voiceless stop pairs are underrepresented, while the nasal-voiceless fricative pairs have O/E-values above 1.

And sonority/nasality:

(26) a. The voiced stop-voiceless fricative pairs are underrepresented though the nasal-voiceless fricative pairs are not.
b. The oral stop pairs occur somewhat less freely than the nasal-voiceless stop pairs.

Here I offer a new account of the influence of the subsidiary features in Muna that makes use of faithfulness constraints. First, we can note that the identical pairs are highly overrepresented. The coronal figures omit the liquids, which are underrepresented.

Labial			**Coronal**			**Dorsal**		
103	59.8	**1.72**	155	106.6	**1.45**	84	64.9	**1.29**

Table 3 Identical consonants in Muna (O, E, **O/E**)

This suggests that the ban against non-identical homorganic consonants may be met (diachronically) through assimilation (see also Rose and Walker 2004 on assimilation that is ‘parasitic’ on place). I adopt the following constraints:

(27)	DOR-ALL	If a sequence of consonants has the same specification for dorsal, then it has the same specification for all features
	LAB-ALL	If a sequence of consonants has the same specification for labial, then it has the same specification for all features
	PLACE-ALL	If a sequence of consonants has the same specification for place, then it has the same specification for all features
	IDENT-VOICE	A segment has the same [voice] specification in Input and Output
	IDENT-SON	A segment has the same [sonorant] specification in Input and Output
	IDENT-CONT	A segment has the same [continuant] specification in Input and Output

I ran a learning simulation with the pairs in Table 2 distributed according to their observed lexical frequency, using the Praat implementation of the GLA, with markedness constraints with an initial value of 100, and faithfulness constraints at 25, and the following settings:

(28) a. Parsing mode set to ‘LinearOT’ (as in the HG model described above)

b. ‘Plasticity’ set at 0.1 (as in Boersma and Levelt 2000)

c. Evaluation noise turned off (this makes it more like Perceptron)

d. "Weighted all" promotion/demotion: The same as the standard GLA ("symmetric all"), except that the amount of ranking change is divided by the number of moving constraints (this helps keep the values above zero)

Acceptability was calculated as in (21) from the H score of the faithful candidate (which violates markedness), and the alternative candidate in which the consonants were identical (which violates faithfulness). The correlation with O/E is very high (*r* = 0.92).

Pair	*Acceptability*	*O/E*	*Pair*	*Acceptability*	*O/E*
gŋ	-20.69	0.00	ts	72.95	0.37
gʁ	-9.23	0.00	kʁ	80.95	0.40
bm	0.13	0.07	td	87.89	0.60
kg	5.71	0.07	pm	90.30	0.39
pf	11.59	0.22	bf	101.77	0.58
pb	26.53	0.10	tn	151.66	0.70
ŋʁ	54.54	0.00	ds	163.12	0.55
dn	61.48	0.25	fm	165.54	1.04
kŋ	69.48	0.10	sn	226.89	1.17

Table 4: Acceptability scores calculated after gradual learning, with O/E

(29) An 'impossible' sequence

Weights	82.17	61.35	2.29	90.18	75.24	63.77	H
/ŋaga/	DOR-ALL	LAB-ALL	PLACE-ALL	IDENT-VCE	IDENT-CONT	IDENT-SON	
[ŋaga]	*		*				84.46
☞ [gaga]						*	63.77

Acceptability([ŋaga]) = 84.46 - 63.77 = **-20.69** (O/E = 0)

Here is a set of tableaux illustrating the account of gradient well-formedness - the difference between the acceptability scores based on the first two tableaux is due to the violations of markedness constraints by the optimal forms, while the differences between the second tableaux and the last two is due to the violations of the faithfulness constraints by the sub-optimal competitors.

(30) *Allowable sequences with different acceptability scores*

Weights	82.17	61.35	2.29	90.18	75.24	63.77	H
/kaga/	DOR-ALL	LAB-ALL	PLACE-ALL	IDENT-VCE	IDENT-CONT	IDENT-SON	
☞ [kaga]	*		*				84.46
[gaga]				*			90.18

Acceptability([kaga]) = 90.18 - 84.46 = **5.71** (O/E = 0.07)

Weights	82.17	61.35	2.29	90.18	75.24	63.77	H
/tada/	DOR-ALL	LAB-ALL	PLACE-ALL	IDENT-VCE	IDENT-CONT	IDENT-SON	
☞ [tada]			*				2.29
[dada]				*			90.18

Acceptability([tada]) = 90.18 - 2.29 = **87.89** (O/E = 0.60)

Weights	82.17	61.35	2.29	90.18	75.24	63.77	H
/tasa/	DOR-ALL	LAB-ALL	PLACE-ALL	IDENT-VCE	IDENT-CONT	IDENT-SON	
☞ [tasa]			*				2.29
[sasa]					*		75.24

Acceptability([tasa]) = 75.24 - 2.29 = **72.95** (O/E = 0.37)

Weights	82.17	61.35	2.29	90.18	75.24	63.77	H
/nada/	DOR-ALL	LAB-ALL	PLACE-ALL	IDENT-VCE	IDENT-CONT	IDENT-SON	
☞ [nada]			*				2.29
[dada]						*	63.77

Acceptability([nada]) = 63.77 - 2.29 = **61.48** (O/E = 0.25)

It is impossible to perfectly match O/E and acceptability with these constraints, since they generalize over individual sequences. Such generalization leads to interesting predictions of differential acceptability for sequences with identical statistical characteristics:

(31) *Zero frequency sequences*

Weights	82.17	61.35	2.29	90.18	75.24	63.77	H
/ŋaga/	DOR-ALL	LAB-ALL	PLACE-ALL	IDENT-VCE	IDENT-CONT	IDENT-SON	
[ŋaga]	*		*				84.46
☞ [gaga]						*	63.77

Acceptability([ŋaga]) = 84.46 - 63.77 = **-20.69** (O/E = 0)

Weights	82.17	61.35	2.29	90.18	75.24	63.77	H
/gaʁa/	DOR-ALL	LAB-ALL	PLACE-ALL	IDENT-VCE	IDENT-CONT	IDENT-SON	
☞ [gaʁa]	*		*				84.46
[gaga]					*		75.24

Acceptability([[gaʁa]]) = 75.24 - 84.36 = **-9.23** (O/E = 0)

Weights	82.17	61.35	2.29	90.18	75.24	63.77	H
/ŋaʁa/	DOR-ALL	LAB-ALL	PLACE-ALL	IDENT-VCE	IDENT-CONT	IDENT-SON	
☞ [ŋaʁa]	*		*				84.46
[ŋaŋa]					*	*	139.01

Acceptability([[ŋaʁa]]) = 139.01 - 84.36 = **54.54** (O/E = 0)

4. Comparisons with other accounts of gradient phonotactics

Difference from a segmental frequency account (e.g. transitional probabilities, O/E; see Albright 2006, Hayes and Wilson *to appear* for related discussion):

(32) Generalization based on phonological constraints leads to predicted asymmetries in structures of equal frequency, based on other properties of the language, and universals that are encoded in the constraint set

Differences from Coetzee's (2004) OT-based rank ordering account:

(33) i. Separates grammatical from ungrammatical forms in the rank ordering (see (19) above)
ii. Connects frequency with relative well-formedness
iii. Generates numerical well-formedness scores whose correlation to lexical statistics and experimental data can be straightforwardly measured

Differences from Frisch *et al.*'s (2004) account of Arabic (Coetzee and Pater 2007):

(34) i. Accounts for cross-linguistic variation in restrictions on homorganic consonants as differences in weighting of constraints
ii. Acceptability scores calculated as above correlate far better with Arabic and Muna O/E scores than do values output by Frisch *et al.*'s similarity metric
iii. Cast in a theory that provides an explicit account of how constraints interact to produce the phonotactics and alternations of the language

Differences from Hayes and Wilson's (to appear) Maximum Entropy model:

(35) i. Allows for constraint conflict; Hayes and Wilson's system seeks to assign a score of '0' to acceptable forms, and thus favors constraints that are unviolated
ii. Does not require the learner to make explicit calculations of frequency or probability

Difference from Frisch et al. (2004), Albright (2006), and Hayes and Wilson (to appear):

(36) Allows a connection between phonotactics and alternations - in this respect, the theory here works in exactly the same way as OT

5. Directions for further research

Genesis of constraints:

(37) Phonotactic knowledge likely involves both universal and language-specific constraints; can this model be supplemented with a mechanism for constraint building?

Relationship to experimental data of phonotactic well-formedness:

(38) i. Modeling existing data
ii. Testing hypotheses about generalization

Relationship to Arabic:

(39) An account of 'OCP-Place subsidiary features' in terms of faithfulness will likely not extend to Arabic, since Arabic bans identical consonants in some positions (see Coetzee and Pater 2007 for a markedness-based approach)

Relationship to learnability theory:

(40) Boersma's GLA has been shown to be non-convergent (Pater to appear); the Perceptron model has convergence proof, but remains to be implemented for Harmonic Grammar, and extended to learning patterns of variation (news on this soon, hopefully!)

Relationship to variation - is the following the right model - and is this compatible with the account of gradient phonotactics above? Or is Boersma's (1998) noise model, or some other theory, better?

(41) *Maximum Entropy model of variation* (Goldwater and Johnson 2003, Jaeger to appear)
Probability of a candidate is equal to the exponent of the sum of its weighted violation scores divided by the sum of same figure for the entire candidate set

Relationship to acquisition data:

(42) i. What level of phonological development is being modeled in the above? Perception, production, the development of the lexicon?
ii. Do stages of acquisition show the cumulative effects of HG constraint interaction? See Pater, Jesney and Tessier (to appear) for some initial results

Relationship to phonological theory:

(43) Is Harmonic Grammar overly powerful as a theory of typology, as claimed by Prince and Smolensky (1993/2004) and Legendre, Sorace and Smolensky (2006), or is it in fact sufficiently restrictive, as claimed by Pater (2007) (see also Pater, Potts and Bhatt 2006)

References

Albright, Adam. 2006. Gradient phonotactic effects: lexical? grammatical? both? neither? LSA talk handout, Jan 7, Albuquerque. (available at http://web.mit.edu/albright/www/)

Boersma, Paul. 1998. Functional Phonology: Formalizing the Interactions between Articulatory and Perceptual Drives, Ph.D. dissertation, University of Amsterdam.

Boersma, Paul, and Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32, 45-86.

Boersma, Paul and Clara C. Levelt. 2000. Gradual constraint-ranking learning algorithm predicts acquisition order. In The proceedings of the thirtieth annual child language research forum, ed. Eve V. Clark. Stanford: CSLI Publications.

Boersma, Paul. 2004. A stochastic OT account of paralinguistic tasks such as grammaticality and prototypicality judgments. Ms., University of Amsterdam. [ROA-648].

Boersma, Paul and David Weenink. 2006. Praat software version 4.4.24.

Coetzee, Andries. 2004. *What it Means to be a Loser: Non-Optimal Candidates in Optimality Theory*. Ph.D. Dissertation, University of Massachusetts.

Coetzee, Andries, and Joe Pater. 2007. Weighted Constraints and Gradient Phonotactics in Muna and Arabic. Ms, University of Michigan and University of Massachusetts, Amherst.

Flemming, Edward. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18, 1.

Frisch, Stefan, Janet Pierrehumbert and Michael Broe. 2004. 'Similarity Avoidance and the OCP', Natural Language and Linguistic Theory 22, 179-228.

Goldwater, Sharon, and Mark Johnson. 2003. Learning OT Constraint Rankings Using a Maximum Entropy Model. In *Proceedings of the Workshop on Variation within Optimality Theory*, ed. Jennifer Spenader, Anders Eriksson, and Östen Dahl, 111-120. Stockholm University.

Hayes, Bruce and Colin Wilson. To appear. A Maximum Entropy Model of Phonotactics and Phonotactic Learning. In *Linguistic Inquiry*.

Jäger, Gerhard. To appear. In *Maximum entropy models and Stochastic Optimality Theory*, ed. Jane Grimshaw, Joan Maling, Chris Manning, Jane Simpson, and Annie Zaenen. Stanford, CA: CSLI. [ROA-625].

Keller, Frank. 2006. Linear Optimality Theory as a Model of Gradience in Grammar. In Fanselow et al. eds., Gradience in Grammar: Generative Perspectives. Oxford University Press.

Legendre, Géraldine, Yoshiro Miyata, and Paul Smolensky. 1990. Harmonic Grammar – a formal multi-level connectionist theory of linguistic wellformedness: An application. In *Proceedings of the twelfth annual conference of the Cognitive Science Society*, 884–891. Cambridge, MA: Lawrence Erlbaum.

Legendre, Géraldine, Antonella Sorace, and Paul Smolensky. 2006. The Optimality Theory–Harmonic Grammar connection. In Smolensky and Legendre (2006), 903–966.

Legendre, Géraldine, Yoshiro Miyata, and Paul Smolensky. 2006. The interaction of syntax and semantics: A Harmonic Grammar account of split intransitivity. In Smolensky and Legendre (2006), 417–452.

Pater, Joe. 2007. The power of weighted constraints. Paper presented at Johns Hopkins University, February 2007. (Available at http://people.umass.edu/pater/)

Pater, Joe. To appear. Non-convergence in the Gradual Learning Algorithm. In *Linguistic Inquiry*.

Pater, Joe, Karen Jesney, and Anne-Michelle Tessier. To appear. Phonological acquisition as weighted constraint interaction. In Proceedings of GALANA 2 (Available at http://people.umass.edu/pater/)

Pater, Joe, Christopher Potts, and Rajesh Bhatt. 2006. Harmonic Grammar with Linear Programming. Ms, University of Massachusetts, Amherst. [ROA-872]

Prince, Alan. 2002. Anything Goes. In New century of phonology and phonological theory, ed. Takeru Honma, Masao Okazaki, Toshiyuki Tabata, and Shin ichi Tanaka, 66–90. Tokyo: Kaitakusha. ROA-536.

Rose, Sharon and Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language* 80, 475-531.

Rosenblatt, Frank. 1962. *Principles of neurodynamics*. Washington, DC. 111-116.

Smolensky, Paul, and Geraldine Legendre. 2006. *The harmonic mind: From neural computation to Optimality-Theoretic grammar*. Cambridge, MA: MIT Press.

van den Berg, René. 1989. A grammar of the Muna language, Foris, Dordrecht.

Weijer, Joost van de 1999. Language input for word discovery. Ph.D. dissertation, Nijmegen University, Nijmegen, The Netherlands.