

Additive Optimization and Phonological Typology*

Joe Pater, UMass Amherst

Research question:

- (1) How do the typological predictions of a theory with weighted constraints and additive evaluation (Additive Optimization) differ from those of standard Optimality Theory?

Main findings to date:

- (2)
 - a. Elimination of gradient alignment constraints eliminates obviously problematic cases produced by Additive Optimization (extending McCarthy 2003a)
 - b. Additive Optimization with categorical constraints does produce gang effects that distinguish it from standard OT (extending Prince 2002)
 - c. Additive Optimization possesses very little of the power of Local Conjunction (extending Legendre, Sorace and Smolensky 2005)

Future research:

- (3) Use of weighting implications to computationally study typology, and for learnability

1. What is Additive Optimization?

OT sets itself apart from other theories of constraint interaction, including Harmonic Grammar (Smolensky *et al.* 1993), in having constraint rankings rather than weightings. When one constraint ranks above others, its preferences are decisive, regardless of the extent of violation of lower ranked constraints.

I use the term Additive Optimization (AO) to refer to a type of theory that replaces OT's constraint ranking with weighting, and evaluates candidates by summing weighted violations. Prince (2002) discusses a simplified version of Harmonic Grammar that does exactly that: the candidate with the lowest sum of weighted violations is optimal (see also Keller 2005). In Harmonic Grammar (see esp. Smolensky and Legendre 2005) the optimal output is one with highest sum of negative weighted violations. See also Flemming (2001), Goldwater and Johnson (2003), Hayes (2005), and Wilson (2005) for phonological applications of AO.

* Thanks to participants in Ling 730, UMass Fall 2005, for their help in developing this material: Paula Aden, Leah Bateman, Michael Becker, Tim Beechey, Mike Key, Anne-Michelle Tessier, and especially Kathryn Flack and Shigeto Kawahara for creating the first restrictiveness proofs using weighting implications. Special thanks also to John McCarthy for sharing an e-mail exchange with Alan Prince on similar issues, and to Paul Smolensky for sharing a pre-publication version of Legendre *et al.* (2005), and to both for extremely useful discussion (and to Sharon Peperkamp for facilitating the discussion with Smolensky). I am also indebted to Rajesh Bhatt, Edward Flemming, Heather Goad, Bruce Hayes, Mark Johnson, Chris Potts, Alan Prince, Doug Pulleyblank and Jason Riggle for further helpful discussion. All misconceptions and errors that remain after all this help are mine alone.

2. Gradiency and Additive Optimization

Legendre, Sorace and Smolensky (2005) show how the interaction of the following two constraints produces distinct results under strict domination and AO:

- (4) **ALIGN-HEAD-R**
 Assess one violation mark for every syllable intervening between the main stress and the right edge of the word
- WEIGHT-TO-STRESS**
 Assess one violation mark for every unstressed heavy syllable

They point out that AO can produce a system that stresses an initial heavy syllable that is followed by three light syllables, but that stresses the final syllable if more syllables follow the initial heavy. Here is a weighting schema that produces this effect, using Prince's (2002) notation in which violation marks are replaced by integers, which are multiplied by weights, and summed:

(5) A problem for Additive Optimization

Weight:	3.5	1	Σ
/bantana/	W-S	ALIGN-HEAD-R	
ban.ta.na.má	1	0	3.5
☞ bán.ta.na.ma	0	3	3

Weight:	3.5	1	Σ
/bantana/	W-S	ALIGN-HEAD-R	
☞ ban.ta.na.va.má	1	0	3.5
bán.ta.na.va.ma	0	4	4

However, if following McCarthy (2003a), the constraint set is purged of gradient constraints like ALIGN-HEAD-R, then this problem vanishes. The violation profiles for the two cases in (5) would be identical, and weighting could not produce distinct results for stress placement.

Gradient constraints, in McCarthy's (2003a) sense, are ones that assess multiple violation marks to a single locus of violation. Multiple violations of categorical constraints do not in general allow for gang effects, because the effect of each violation on the outcome of evaluation is independent. Consider the following two tables:

(6) Multiple violations of categorical constraints

Weight:			Σ
/batan/	NOCODA	MAX	
ba.tan	1	0	
ba.ta	0	1	

Weight:			Σ
/bantana/	NOCODA	MAX	
ban.tan	2	0	
ba.ta	0	2	

The number of codas does not affect the outcome: if $w(\text{NOCODA}) > w(\text{MAX})$, deletion applies in both cases, and if $w(\text{MAX}) > w(\text{NOCODA})$, the faithful candidate wins both times. There is no way for an increase in the number of codas to force deletion.

Gradient constraints allow an *asymmetric trade-off* in constraint violations between candidates, which is necessary to produce gang effects. In (5), multiple violations of ALIGN-HEAD-R trade against a single violation of W-S. In (6), however, NOCODA and MAX trade off one-to-one.

3. Gang Effects with Categorical Constraints

Under the right circumstances, categorical constraints can produce asymmetric trade-offs. For this to happen with just a pair of constraints, a single violation of one constraint must be enough to avoid violations of another constraint in two or more separate loci. Prince (2002: 19) identifies one such case for MAX and DEP; one epenthetic vowel trades for two consonant deletions.

(7) Two deletions trade for one insertion

/CC/	DEP	MAX
∅	0	2
CVC	1	0

Prince (2002) does not pursue the typological implications of the AO interaction. It opens up the possibility that a single underlying consonant could be deleted to meet syllable structure requirements, while a pair of consonants could be syllabified with a vowel:

(8) Trade-off instantiated with *COMPLEX-CODA

Weight:	1.5	1	2	Σ
/VCCC/	DEP	MAX	*COMP-CODA	
VC	0	2	0	2
VC.CVC	1	0	0	1.5
VCCC	0	0	1	2

Weight:	1.5	1	2	Σ
/VCC/	DEP	MAX	*COMP-CODA	
VC	0	1	0	1
V.CVC	1	0	0	1.5
VCC	0	0	1	2

Unlike the gradient alignment example, this is not an obviously pathological system (nor is it obviously attested). It is difficult to construct further cases of pairwise asymmetric trade-offs, except for ones involving the Null Parse, which is a single global violation that can be traded against anything (a good topic for our discussion period, in light of “Wolf today”)

Violations of multiple constraints can also be traded off against each other. For example, a pair of markedness constraints can gang up on a single faithfulness constraint, as illustrated in (9) with a case from Japanese loanword phonology (thanks to Shigeto Kawahara for this example). Geminate (optionally) devoice in the presence of another voiced segment (Nishimura 2003, Kawahara to appear):

(9) Cumulative effect of OCP-VOICE and *VCE-GEM

Weight	3	2	2	Σ
/beddo/ 'bed'	ID-VCE	OCP-VOICE	*VCE-GEM	
[beddo]		1	1	4
[betto]	1			3

Weight	3	2	2	Σ
/bagii/ 'buggy'	ID-VCE	OCP-VOICE	*VCE-GEM	
[bagii]		1		2
[bakii]	1			3

<i>Weight</i>	3	2	2	Σ
/reddo/ 'red'	ID- VCE	OCP- VOICE	*VCE- GEM	
ر [reddo]			1	2
[retto]	1			3

Such gang effects can be modeled in standard OT by increasing the size of the constraint set (see Nishimura 2003 for a Local Conjunction analysis of this case, and Kawahara to appear for one employing geminate-specific faithfulness). However, given the set of constraints in (9), no ranking will produce this pattern.¹

Additive interactions of markedness and faithfulness constraints are also possible. One attested pattern this produces is what McCarthy (2003b) refers to as a Grandfather effect: structures present in the Input can violate a markedness constraint, but derived structures are forced to obey it, even at the cost of violating another active markedness constraint. This is illustrated in (10) with McCarthy's Meccan Arabic example (Bakalla 1973, Abu-Mansour 1996), in which voiced obstruents are banned as the output of voicing assimilation, but not otherwise.

(10) **A grandfather effect as additivity**

<i>Weight</i>	5	4	2	Σ
/ʔakbar/	AGR- VCE	ID- VCE	*VCE	
[ʔagbar]		1	2	8
ك [ʔakbar]	1		1	7

<i>Weight</i>	5	4	2	Σ
/ʔagsam/	AGR- VCE	ID- VCE	*VCE	
[ʔagsam]	1		1	7
ك [ʔaksam]		1		4

<i>Weight</i>	5	4	2	Σ
/ʔibnu/	AGR- VCE	ID- VCE	*VCE	
ب [ʔibnu]			1	2
[ʔipnu]		1		4

Here violations of IDENT-VOICE and *VOICE trade off against a single violation of AGREE-VOICE. This analysis is similar to the Local Conjunction analysis of Grandfather Effects found in Bakovic (2000) and Smolensky (2005), though as we will see, its typological predictions are different. See McCarthy (2002) for discussion of further Grandfather Effects, including a case that escapes both Local Conjunction and AO (though see de Lacy 2002 for an analysis of Sri Lankan Portuguese Creole in standard OT).

¹ As Bruce Hayes (p.c.) points out, additive effects are possible in stochastic OT (Boersma 1998, Boersma and Hayes 2001). If two constraints probabilistically outrank one conflicting constraint, then a form in which both of those constraints conflict with the single constraint is more likely to violate it. However, this type of additivity will not capture either of the cases discussed here, both of which involve categorical distinctions (see Kawahara to appear on this aspect of the Japanese example).

4. Additive Optimization and Local Conjunction

So far we have seen examples of attested systems produced by AO with categorical constraints. Does AO also continue to produce improbable phonologies if we eliminate gradient constraints?

In an attempt to construct such systems, I tried to replicate pathological Local Conjunctions using AO (see Smolensky 1995, 2005 on local conjunction; see Itô and Mester 1998, McCarthy 1999, 2002, 2003b, Padgett 2002, Kawahara to appear on its pathologies). However, the asymmetric trade-off requirement robs AO of most of the power of Local Conjunction. Local Conjunction produces *super-additive* interactions (Legendre *et al.* 2005), which do not require asymmetric trade-offs.

Because violations of two constraints in an AO gang effect have to trade off against just a single violation of another constraint, there is an inherent locality restriction that is either absent, or stipulated, in Local Conjunction theory. Consider the self-conjoined NOCODA² in (11), which takes as its domain the PrWd.

(11) Non-local Local Conjunction

/batan/	NOCODA ²	MAX	NOCODA
☞ ba.tan			*
ba.ta		*!	

/bantān/	NOCODA ²	MAX	NOCODA
ban.tan	*!		**
☞ ban.ta		*	*

As we saw in (6), because the violations of MAX and NOCODA trade off one-to-one, this system cannot be modeled in AO.

Non-local interactions like that in (11) are the source of many of the pathological systems produced by Local Conjunction (see esp. McCarthy 1999, 2003a). They are also potentially the source of some attested ones, like OCP effects (e.g. Alderete 1997, Itô and Mester 1998, Suzuki 1998), but no proposal has successfully distinguished them. Since AO cannot replicate self-conjunction it also cannot replicate the counting pattern discussed in Kawahara (to appear).

Because a pair of constraints in an AO gang effect must both trade-off against a single competing constraint, there is also a “co-relevance” requirement that is again absent or stipulated in Local Conjunction (see again McCarthy 2003a). The following tableaux show how conjoined *V-NASAL and AGREE-BACK can produce backness assimilation only when one of the vowels is nasal (or equally bizarrely, denasalization in the presence of a disharmonic vowel, if the ranking of ID-NAS and ID-BACK is switched). This example is a problematic consequence of the Local Conjunction approach to vowel harmony pursued in Kiparsky and Pajusalu (2003) and Smolensky (2005).

(12) [+nasal] as a target/trigger condition for backness harmony

/idũ/	*V-NAS& AGR-BACK	ID- NAS	ID-BACK	*V-NAS	AGR- BACK
☞ [idỹ]			*!	*	
[idũ]	*!			*	*

/idu/	*V-NAS& AGR-BACK	ID- NAS	ID-BACK	*V-NAS	AGR- BACK
[idy]			*!		
☞ [idu]					*

The annotated mappings in (13) show that the AO weighting conditions needed for this system are inconsistent, if the constraint set includes only the unconjoined constraints from (12). Since the *V-NAS violation is shared by both candidates in (13a), it is correctly irrelevant.

- (13) a. /idũ/ → [idỹ] *[idũ] $w(*V-NAS) + w(AGR-BACK) > w(ID-BACK) + w(*V-NAS)$
b. /idu/ → [idu] *[idy] $w(ID-BACK) > w(AGR-BACK)$

Conjoined markedness and faithfulness constraints can also produce implausible results. In Itô and Mester's (1998) example, a faithfulness constraint protects a structure from the effects of a markedness constraint only when a second markedness constraint is violated (thanks to Matt Wolf for pointing out the relevance of this case):

(14) **Preservation of a feature only in a marked environment**

/ad/	NoCODA& IDENT-VCE	*VCE	ID-VOICE
☞ [ad]		*	
[at]	*!		*

/da/	NoCODA& IDENT-VCE	*VCE	ID-VOICE
[da]			*!
☞ [ta]		*	

This is again a relevance problem: NOCODA does not conflict with *VCE, but this is not a requirement for the super-additive interactions produced by Local Conjunction. In AO, however, the NOCODA violation in [ad] is correctly irrelevant to the decision of whether to preserve voicing. This is again demonstrated in terms of inconsistent weighting conditions in (15).

- (15) a. /ad/ → [ad] *[at] $w(*NOCODA) + w(ID-VCE) > w(*VCE) + w(NOCODA)$
b. /da/ → [ta] *[da] $w(*VCE) > w(ID-VCE)$

An important caveat about these restrictiveness results for AO: as usual, they are dependent on the content of the constraint set. In particular, if segmental deletion violates MAX, but not featural faithfulness (i.e. IDENT), as in standard correspondence theory, then a pair of quite unrelated markedness constraints could gang up (substitute MAX for the two IDENT constraints in (12)). However, both constraints have to be satisfied by deletion; this may be key to ruling out some cases (e.g. if deletion does not satisfy AGREE). And again, if the Null Parse violates no other faithfulness constraints, then it can be forced by any pair of markedness constraints.

Legendre *et al.* (2005) argue that the weakness of AO relative to Local Conjunction is to its detriment. Their example is an attested markedness conjunction (English *[x] as *DORSAL&*[+cont]) that has the same characteristics as that in (12); a pair of markedness constraints gang up to overcome two higher ranked faithfulness constraints.

The main goal of Smolensky's (1995) original proposal for Local Conjunction was to derive complex contextual markedness constraints from the conjunction of more basic constraints (e.g. *CODA[VOICE] as *CODA&*[VOICE]). However, the set of contextual markedness constraints that results from conjunctions of context-free markedness constraints with each other and with prosodic markedness constraints does not seem to accurately reflect typology (cf. e.g. *NC_o, Steriade 2001). Since the content of this set of constraints appears to be substantively delimited, it is not clear what is gained from decomposition (see Padgett 2002 for related discussion). Therefore, it is not clear that we need (or want) AO to mimic either *[x] or *CODA&*[VOICE].

Local Conjunction is extended to conjunctions of faithfulness constraints for the analysis of chain shifts (counter-feeding opacity) in Kirchner (1996) and Moreton and Smolensky (2002). AO cannot replicate this analysis, as shown schematically in (16), in which the columns M and F indicate the markedness and faithfulness constraints violated by each mapping (this proof is originally due to Kathryn Flack). The Local Conjunction analysis would block A->C with the conjoined F1&F2.

(16)		M	F	<i>Inconsistency under AO</i>
	A -> A	*A		
	☞ A -> B	*B	F1	Therefore w(F1) +w(F2) > w(*B)+ w(F1)
	A -> C		F1+F2	
	B -> B	*B		
	☞ B -> C		F2	Therefore w(*B) > w(F2)

However, as McCarthy (2003b) argues, conjunction of faithfulness constraints is also subject to locality and relevancy problems, so another approach to counter-feeding opacity is required.

We have seen that AO can mimic the local conjunction analysis of Grandfather Effects. Lubowicz (2002) proposes another use of markedness/faithfulness conjunction: to capture Derived Environment Effects. In the schematic example in (17), conjoined *B&F1 would block A->B. This again escapes AO, as shown by the inconsistent weighting implications.

(17)		M	F	<i>Inconsistency under AO</i>
	A -> A	*A		
	A -> B	*B	F1	
	☞ A -> C		F1+F2	Therefore w(*B)+ w(F1) > w(F1) +w(F2)
	☞ B -> B	*B		Therefore w(F2) > w(*B)
	B -> C		F2	

See again McCarthy (2003b) for reasons to be cautious about the Local Conjunction analysis.

5. Toward the computational study of the typology of AO

Determining the extent to which additive interactions are welcome or not in phonological theory will require considerable further study of the typology produced by AO. A better understanding of the power of AO is also necessary to more precisely specify the restrictions imposed by strict domination.

One barrier to systematically studying the typology of AO is that it is not simple to infer from a set of candidate Input-Output mappings, constraints, and violation marks, what the possible sets of I-O mappings are. Even for standard OT, which has more restricted interactional possibilities, a computational tool for the calculation of factorial typology is extremely useful for constraint sets of any size (OTSoft, Hayes *et al.* 2003).

Above weighting implications were used to show that some I-O mappings cannot co-exist in a single language, given a particular constraint set. In collaboration with Rajesh Bhatt, I am pursuing the possibility of using such implications to computationally investigate the typological predictions of AO. The basic idea is simple: the algorithm would determine which sets of optimal outputs can co-exist by determining whether their weighting implications are consistent. That is, given a set of candidate I-O mappings as in (18a) along with associated constraints and violation marks, the question is which of the 27 sets of optima in (18b) would display consistent weighting requirements.

(18) Schematization of the search space

a.	Input 1	Can1a Can1b Can1c	Input 2	Can2a Can2b Can2c	Input 3	Can3a Can3b Can3c
----	---------	-------------------------	---------	-------------------------	---------	-------------------------

b. {Can1a, Can2a, Can3a}, {Can1a, Can2a, Can3b}, {Can1a, Can2a, Can3c}, {Can1a, Can2b, Can3a}, {Can1a, Can2b, Can3b}, {Can1a, Can2b, Can3c}, {Can1a, Can2c, Can3a}, {Can1a, Can2c, Can3b}, {Can1a, Can2c, Can3c}, {Can1b, Can2a, Can3a}, {Can1b, Can2a, Can3b}, {Can1b, Can2a, Can3c}, {Can1b, Can2b, Can3a}, {Can1b, Can2b, Can3b}, {Can1b, Can2b, Can3c}, {Can1b, Can2c, Can3a}, {Can1b, Can2c, Can3b}, {Can1b, Can2c, Can3c}, {Can1c, Can2a, Can3a}, {Can1c, Can2a, Can3b}, {Can1c, Can2a, Can3c}, {Can1c, Can2b, Can3a}, {Can1c, Can2b, Can3b}, {Can1c, Can2b, Can3c}, {Can1c, Can2c, Can3a}, {Can1c, Can2c, Can3b}, {Can1c, Can2c, Can3c}

As (18b) illustrates, the search space of sets of optima will be extremely large. However, there are almost certainly more efficient ways to find the set of possible I-O mappings than to generate all of the sets and check them one-by-one.

Each of the members of one of these sets of optima has a set of entailed weighting conditions, as the shown on the next page.

The following table schematizes the information given for a single Input with three candidate Outputs, and three constraints. ‘CXY’ represents the number of violations marks incurred candidate X on constraint Y.

(19) **Information for one Input**

	w_1	w_2	w_3
/Input1/			
Cand1	C11	C12	C13
Cand2	C21	C22	C23
Cand3	C31	C32	C33

If Cand1 is the optimal candidate, then the following implications hold:

$$(20) \quad \begin{aligned} w_1(C21) + w_2(C22) + w_3(C23) &> w_1(C11) + w_2(C12) + w_3(C13) \\ w_1(C31) + w_2(C32) + w_3(C33) &> w_1(C11) + w_2(C12) + w_3(C13) \end{aligned}$$

Which can be restated as the following equations:

$$(21) \quad \begin{aligned} w_1(C21-C11) + w_2(C22-C12) + w_3(C23-C13) &> 0 \\ w_1(C31-C11) + w_2(C32-C12) + w_3(C33-C13) &> 0 \end{aligned}$$

It should be possible to apply what is known as the simplex method in linear programming to determine whether equations of the type in (21) are consistent or not.

Inconsistency amongst the equations (21) would imply that C1 is harmonically bounded (see Prince 2002 for discussion of harmonic bounding in AO).

Inconsistency amongst equations from two optima (from different inputs) indicates that a language including both of them is impossible.

Calculation of typology via weighting implications is probably more feasible than to translate the OT method of factorial typology into AO, given that the set of possible weightings for even a pair of constraints is infinite.

Weighting implications might also be fruitfully studied by a learner. An important property of the Constraint Demotion Algorithm is that it detects inconsistency; this has been applied to the learning of various types of hidden structure (Tesar 1998 *et seq.*).

Most implementations of AO use learning algorithms that have a weaker criterion for success than the Constraint Demotion Algorithm: they seek to optimize the fit of the weightings to the data, rather than ensure that the weightings do generate the data. Equipping a learner of an AO system with the ability to detect inconsistencies might also allow these learners to recognize faulty assumptions about hidden structure (though this might not be necessary if these assumptions are themselves probabilistic). It could also be possible to build a learner that uses nothing but the implications to infer weightings.

6. Conclusions and questions

The typological predictions of AO are at present understudied, perhaps because they are thought to be obviously out of line with empirical observations, and/or because they are thought to be too difficult to assess. There is reason to believe neither of these views is accurate.

A serious study of the typology of AO will have to address a number of questions. Here are some of them:

- (22)
- a. To what extent are predicted gang effects observed?
 - b. Is the categorical vs. gradient distinction the right cut to make between constraints that are in and out of AO's Con? See Flemming (2001) for a theory of AO that makes use of gradient constraints, but ones quite different in character from gradient Alignment constraints.
 - c. Is simple addition the right mode of interaction?
 - d. What about ineffability? The Null Parse doesn't seem amenable to AO.
 - e. Unlike standard OT, AO could potentially allow for "positive constraints", constraints that reward a candidate for possessing a particular structure (Mark Johnson, p.c.). What would the consequences be?

References

- Abu-Mansour, Mahasen Hasan. 1996. Voice as a privative feature. *Perspectives on Arabic Linguistics VIII: Papers from the Eighth Annual Symposium on Arabic Linguistics*. Ed. Mushira Eid. Amsterdam and Philadelphia: John Benjamins.
- Alderete, John. 1997. Dissimilation as local conjunction. In *NELS 27: Proceedings of the North East Linguistic Society 27*, ed. Kiyomi Kusumoto, 17-32. Amherst, MA: GLSA Publications.
- Bakalla, Mohammed. 1973. *The Morphology and Phonology of Meccan Arabic*. Doctoral dissertation. School of Oriental and African Studies, University of London.
- De Lacy, Paul. 2002. *The formal expression of markedness*. PhD dissertation. University of Massachusetts, Amherst. ROA. 542.
- 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.
- Flemming, Edward. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18. 7-44.
- Goldwater, S., and M. Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. *Proceedings of Stockholm Workshop on Variation within Optimality Theory*.
- Hayes, B. 2005. A generative model of phonotactics. Lecture handout, University of Michigan.

- Hayes, Bruce, Bruce Tesar, and Kie Zuraw (2003) "OTSoft 2.1," software package, <http://www.linguistics.ucla.edu/people/hayes/otsoft/>
- Itô, Junko, and Armin Mester. 1998. *Markedness and Word Structure: OCP Effects in Japanese*. Ms, University of California at Santa Cruz.
- Itô, Junko, and Armin Mester. 2003. *Japanese Morphophonemics: Markedness and Word Structure*. MIT Press.
- Kawahara, Shigeto. To appear. A faithfulness ranking projected from a perceptibility scale: The case of [+voice] in Japanese. In *Language*.
- Keller, F. 2005. Linear Optimality Theory as a Model of Gradience in Grammar. In Fanselow *et al.* eds., *Gradience in Grammar: Generative Perspectives*. Oxford University Press.
- Kiparsky, P. and K. Pajusalu. 2003. Towards a Typology of Disharmony. *The Linguistic Review* 20. 217-241.
- Legendre, G., A. Sorace and P. Smolensky. 2005. The Optimality Theory – Harmonic Grammar Connection. In Smolensky and Legendre *eds.*
- McCarthy, J. 1999. Sympathy and Phonological Opacity. *Phonology* 16. 331-399.
- McCarthy, J. 2002. Comparative markedness [long version]. In Angela C. Carpenter, Andries W. Coetzee, and Paul de Lacy, eds., *Papers in Optimality Theory II* (University of Massachusetts Occasional Papers in Linguistics 26). Amherst, MA: GLSA. Pp. 171-246.
- McCarthy, J. 2003a. OT constraints are categorical. *Phonology* 20, 75-138.
- McCarthy, J. 2003b. Comparative markedness [short version]. *Theoretical Linguistics* 29. 1-51.
- Nishimura, K. 2003. Lyman's law in loanwords. Ms, University of Tokyo.
- Padgett, J. 2002. Constraint Conjunction versus Grounded Constraint Subhierarchies in Optimality Theory. Ms, University of California, Santa Cruz
- Prince, A. 2002. Anything Goes. Ms, Rutgers University.
- Smolensky, P. 1995. On the internal structure of the constraint component Con of UG. Handout of talk given at University of Arizona, March 1995.
- Smolensky, P. 2005. Optimality in Phonology II: Harmonic Completeness, Local Constraint Conjunction, and Feature Domain Markedness. In Smolensky and Legendre (eds.)
- Smolensky, P., G. Legendre, and Y. Miyata. 1993. Integrating connectionist and symbolic computation for the theory of language. *Current Science* 64: 381–391.
- Smolensky, P. and G. Legendre (eds.). 2005. *The Harmonic Mind*. MIT Press.
- Suzuki, Keichiro. 1998. *A Typological Investigation of Dissimilation*. Ph.D. dissertation, University of Arizona.
- Tesar, Bruce. 1998. Using the mutual inconsistency of structural descriptions to overcome ambiguity in language learning. In Proceedings of NELS 28, ed. P. Tamanji and K. Kusumoto, 469-483. Amherst, MA: GLSA, University of Massachusetts.
- Wilson, Colin. 2005. Learning Phonology with Substantive Bias: An Experimental and Computational Study of Velar Palatalization. Ms, UCLA.