Cumulative ill-formedness in typological and experimental data

Joe Pater, University of Amherst, Massachusetts

Conference on Experimental Approaches to Optimality Theory, University of Michigan, 5/19/07

1 Cumulative ill-formedness in experimental data

Ohala and Ohala (1986) present experimental results that are problematic for Chomsky and Halle's (1968) model of phonotactic knowledge. They point out that the SPE theory does not distinguish between a word with one ill-formed structure and a word with multiple instances of ill-formedness. Cumulative ill-formedness equally plays no role in standard Optimality Theory (OT: Prince and Smolensky 2004).

In their pilot study, they show that English speakers do distinguish between nonce words on this basis:

Table 13.6
Stimuli and Results of Experiment 3: Greenberg and Jenkins versus Chomsky and Halle (see Table 13.4 for predictions of the two models)

Stimuli		No. subjects	
A	B	A closer to English	B closer to English
mlɔʒ	mlit	4	11[a]
mløf	spøf	1	15
xlox	xrit	5	11
sfub	sfɛt	1	15
Total		11	52*

[a]One of the 16 subjects gave no response.
*$p < .001$.

Though the results are obviously preliminary, the effect does look robust. Perhaps because it seems so clear that it would hold up in a more systematic study, there seems not to have been a subsequent investigation of the effect of cumulativity in word-likeness/acceptability judgments (though see the discussion of Hay et al. 2004 in Pierrehumbert 2001).

2 Absence of cumulativity in typological data

In this section, I show how unrestricted cumulativity produces implausible phonological systems. To do so, I provide examples that use Optimality Theory with Local Conjunction (OT-LC; Smolensky2006: 43); this draws on previous critiques of OT-LC (esp. McCarthy1999; 2003). See Lubowicz (2005) for recent discussion of how imposing limits on conjunction may solve some of these problems, as well as Moreton and Smolensky (2002) for inherent restrictions on OT-LC analyses of chain shifts.

Smolensky (2006: 43) defines local conjunction as in (1).

(1) Local conjunction within a domain *D*

*A&$_D$ *B is violated if and only if a violation of *A and a (distinct) violation of *B both occur within a single domain of type *D*.

The example in (2) shows how a conjunction of two independently motivated constraints can be used to produce an attested system that the unconjoined constraints cannot yield.

(2) Local conjunction analysis of final devoicing

/bad/	NoCoda&$_{Seg}$*VoiceObs	Ident-Vce	NoCoda	*VoiceObs
☞ bat		*	*	*
bad	*!		*	**
pat		**!	*	
pad	*!	*	*	*

The example in (3) shows how a different conjunction yields an unattested pattern (Ito and Mester 1998):

(3) Local conjunction analysis of initial devoicing

/bad/	NoCoda&$_{Seg}$Ident-Voice	*VoiceObs	NoCoda	Ident-Vce
☞ pad		*	*	*
bat	*!	*	*	*
bad		**!	*	
pat	*!		*	**

Unattested systems can also be produced by conjoining markedness constraints. One such system devoices an obstruent if there is a coda anywhere in the word:

(4) a. /batalak/ →[patalak]

b. /batala/ → [batala]

This system emerges if we conjoin NoCoda and *VoiceObs over the domain of the word:

(5) Local conjunction analysis of long-distance cumulativity

/batalak/	NoCoda&$_{Wd}$*VoiceObs	Ident-Vce	NoCoda	*VoiceObs
batalak	*!		*	*
☞ patalak		*	*	

/batala/	NoCoda&$_{Wd}$*VoiceObs	Ident-Vce	NoCoda	*VoiceObs
☞ batala				*
patala		*!		

If we self-conjoin a markedness constraint over the domain of the word, we limit words to one instance of the marked structure (note that recursive self-conjunction allows a limit of 2, of 3, of 4,...):

(6) Local conjunction analysis of long-distance cumulativity with a single constraint

/bat/	NoCoda&$_{Wd}$NoCoda	Max	NoCoda
bat		*!	
☞ bat			*

/bantat/	NoCoda&$_{Wd}$NoCoda	Max	NoCoda
☞ banta		*	*
bantat	*!		**

While some attested phenomena can be analyzed in this way (e.g. Alderete 1997, Suzuki 1998, Itô and Mester 2003), no proposal has successfully distinguished them from the vast range of cases that do not occur.

Markedness constraints conjoined in their smallest common domain can also yield implausible systems. For example, a conjunction of two constraints demanding assimilation can create a system in which assimilation of one feature occurs only when the segments disagree in their specification for the other feature.

(7) Local conjunction analysis of "anti-parasitic assimilation"

/bo+ty/	Agree-Back&Agree-Height	Ident-Back	Agree-Back	Agree-Height
[boty]	*!		*	*
☞[botu]		*		*

/bu+ty/	Agree-Back&Agree-Height	Ident-Back	Agree-Back	Agree-Height
☞[buty]			*	
[butu]		* !		

Attested languages display the reverse pattern of parasitic assimilation (see recently Rose and Walker 2004).

The dilemma:

(8) Acceptability ratings show evidence of cumulative ill-formedness, but unrestricted cumulative constraint interaction produces implausible phonological systems

The proposed resolution:

(9) Harmonic Grammar provides a restrictive theory of cumulative interaction in Input-Output mappings, and can also be used to model the (presumably) less restricted cumulativity in acceptability ratings

3 Cumulativity and its absence in HG I-O mappings

In Harmonic Grammar (HG; Legendre et al. 1990b,c; Smolensky and Legendre 2006), representations are evaluated in terms of their Harmony, which is calculated by a simple linear equation. The representations' scores on a set of constraints $\{C_1, C_2, C_3,...C_n\}$ are each multiplied by the constraint's coefficient, or weight (W), and are then summed:

(10) $\mathcal{H} = C_1W_1 + C_2W_2 + C_3W_3, + ... + C_nW_n$

As Prince and Smolensky (1993/2004: 236) point out, this score can be used to choose the optimal mapping between levels of representation in an OT-like theory of generative grammar.

If we convert violations to corresponding negative integers, and the optimum is the candidate with maximal Harmony, then constraints with higher weights will have stronger effects in determining the optimum:

(11) A weighted constraint tableau

Weight	1.5	1	$\mathcal{H}$
/Input/	Constraint-1	Constraint-2	
☞ Output-1		-1	-1
Output-2	-1		-1.5

Here the outcome is identical to an OT ranking Constraint-1 >> Constraint-2.

The following tableau illustates a scenario in which the result of the HG weighting diverges from the corresponding ranking (a gang effect; termed a groupwise asymmetric trade-off in Pater 2007):

(12) HG ≠OT

Weight	1.5	1	1	$\mathcal{H}$
/Input/	Constraint-1	Constraint-2	Constraint-2	
Output-1		-1	-1	-2
☞ Output-2	-1			-1.5

The pair of tabeaux in (11) and (12) illustrate the greater power of HG relative to OT: there is no ranking of these constraints that will produce these optima. Prince and Smolensky (1993/2004: 236) and Legendre et al. (2006b) fear that this power is excessive.

The main point of this section (see also Legendre et al. 2006b):

(13) In terms of its ability to produce cumulative interactions, HG is (perhaps surprisingly) restrictive

In particular:

(14) It generates none of the cases of cumulative interaction discussed in section 2 because none fit the violation profile in (12)

First, initial devoicing. The desired optimum is shown by a pointing finger. However, no weighting can in fact choose that optimum, since it is harmonically bounded by final devoicing (NoCodaVoice is included because coda devoicing is also impossible without it).

(15) No initial devoicing in HG

/bad/	NoCodaVoice	*VoiceObs	NoCoda	Ident-Vce
☞ pad	-1	-1	-1	-1
bat		-1	-1	-1
bad	-1	-2	-1	
pat			-1	-2

Next, long distance coda-driven devoicing:

(16) An impossible pair of optima in HG

/batalak/	IDENT-VCE	NOCODA	*VOICEOBS
batalak		-1	-1
☞ patalak	-1	-1	

/batala/	IDENT-VCE	NOCODA	*VOICEOBS
☞ batala			-1
patala	-1		

Here there is no weighting that will simultaneously choose both optima. For a candidate to be picked as optimal, its summed weighted violation score must be higher than its competitors. For /batalak/ →[patalak], we thus have the entailed inequality in (17a.), which is equivalent to (17b.)

(17) a. w(NoCoda) + w(*VoiceObs) > w(NoCoda) + w(Ident-Vce)

b. w(*VoiceObs) > w(Ident-Vce)

From /batalak/ → [batala], we can infer the following weighting condition:

(18) w(Ident-Vce) > w(*VoiceObs)

These are obviously inconsistent - the moral:

(19) Optimality is comparative and relative to a given Input: what is important is the difference between candidates.

Because of the comparative nature of optimality, the shared violation of NoCoda in (16) is irrelevant. The same point applies to the "anti-parasitic assimilation" case, and to multiple violations of a single constraint. A system in which codas are eliminated when their number exceeds *n* is impossible in HG:

(20) No one-coda maximum in HG

/bat/	NOCODA	MAX
☞ bat	-1	
ba		-1

/bantat/	NOCODA	MAX
bantat	-2	
☞ banta	-1	-1

The inconsistent weighting conditions:

(21) a. /bat/ → [bat]: w(Max) > w(NoCoda)

b. /bantat/ → [bantat]: w(NoCoda) > w(Max)

The shared NoCoda violation in [banta] and [bantat] is irrelevant.

These results would not hold if we set a numerical cut-off on well-formedness. For example, if we defined well-formed mappings as those with $\mathcal{H}$ > -1.5, we could produce the one-coda maximum, as shown in (22).

(22) Evaluation of mappings with cut-off of -1.5

Weight	1	$\mathcal{H}$
	NoCoda	
√ /bat/ →[bat]	-1	-1
√ /bantat/ →[banta]	-1	-1
* /bantat/ →[bantat]	-2	-2

So what can HG do? Shigeto Kawahara (p.c.) brings the following attested cumulative interaction to my attention (Nishimura2003, Kawahara 2006). It provides a useful example of the conditions that must obtain for HG to generate such an interaction.

In Japanese loanwords, pairs of voiced obstruents are permitted (data from Kawahara 2006):

(23) Violations of Lyman's Law in loanwords

bagii 'buggy'	bagu 'bug'	gibu 'give'
bogii 'bogey'	dagu 'Doug'	
bobu 'Bob'	giga 'giga'	

Voiced obstruent geminates also occur in loanwords:

(24) Voiced/voiceless obstruent geminate near-minimal pairs in Japanese loanwords

webbu 'web'	wippu 'whipped (cream)'
sunobbu 'snob'	sutoppu 'stop'
habburu 'Hubble'	kappuru 'couple'
kiddo 'kid'	kitto 'kit'
reddo 'red'	autoretto 'outlet'
heddo 'head'	metto 'helmet'

When a word contains both a voiced geminate and a voiced obstruent, the geminate is optionally, but categorically, devoiced:

(25) Optional devoicing of a geminate in Lyman's Law environment

guddo ~ gutto 'good'	doggu ~ dokku 'dog'
beddo ~ betto 'bed'	baggu ~ bakku 'bag'
doreddo ~ doretto 'dredlocks'	kappuru 'couple'
baddo ~ batto 'bad'	doraggu ~ dorakku 'drug'
deibiddo ~ deibitto 'David'	biggu ~ bikku 'big'
budda ~ butta 'Buddha'	

Such devoicing is judged unacceptable in both (23) and (24).

In HG, this devoicing pattern can be analyzed as an instance of cumulative interaction between *Voice-Obs and Ito and Mester's (1986) OCP-Voice (see Nishimura's 2003 OT-LC analysis; see Kawahara 2006 for a critique and a different expansion of the OT constraint set):

(26) Japanese loanword devoicing as cumulative constraint interaction

	Weight	1.5	1	$\mathcal{H}$
	/bobu/	IDENT-VOICE	OCP-VOICE	
☞	[bobu]		-1	-1
	[bopu]	-1		-1.5

	Weight	1.5	1	$\mathcal{H}$
	/webbu/	IDENT-VOICE	*VOICE-OBS	
☞	[webbu]		-1	-1
	[weppu]	-1		-1.5

	Weight	1.5	1	1	$\mathcal{H}$
	/doggu/	IDENT-VOICE	*VOICE-OBS	OCP-VOICE	
	[doggu]		-1	-1	-2
☞	[dokku]	-1			-1.5

In the last tableau, violations of *Voice-Obs and OCP-Voice trade-off against a single violation of Ident-Voice.

The requirement that constraint violations trade-off in this way between candidates places tight restrictions on the set of cumulative interactions produced by HG in Input-Output mappings (though see Pater 2006; 2007 for discussion of further issues). The result is a theory of cumulative constraint interaction with inherent restrictions on *co-relevance* and *locality* (cf. McCarthy's 2003 critique of OT-LC).

4 Gradient acceptability in HG and OT

Gradient acceptability can be categorized into two types - here I provide examples from phonotactics:

(27) Types of gradient acceptability

i. *Gradient well-formedness*: Two attested structures in a language differ in their acceptability

Hebrew roots with identical consonants in C2 and C3 (SMM) vs. heterorganic sequences (PSM) (Berent and Shimron 1997; Berent et al. 2001)

Arabic roots with homorganic consonants of different degrees of similarity (Frisch and Zawaydeh 2001)

English singleton onsets and codas of different degrees of sonority (Moreton et al. 2005)

English sKVK vs. sTVT (Coetzee 2004;Pater and Coetzee 2005)

ii. *Gradient ill-formedness*: Two unattested structures differ in their acceptability

Cumulative ill-formedness (see section 1)

Onset clusters unattested in English (Scholes 1966; Pertz and Bever 1975; Smolensky et al. 2003; Berent et al. 2006)

English sKVK vs. sPVP (Coetzee 2004)

I abstract from variation between outcomes for a single Input, which can be seen as a grey area between a structure being attested and unattested (see Boersma and Hayes 2001 on the connection between frequency of variants and well-formedness judgments). I also abstract from possible mismatches between "grammatical-ungrammatical" and "attested-unattested" (see e.g. Moreton 2002).

In this section, I build on Boersma's (2004) critique of Keller's (2000; 2006) HG account of gradient acceptability to argue that:

(28) No extant HG or OT approach to gradient acceptability satisfactorily accounts for both gradient well-formedness and gradient ill-formedness

I also propose a new metric for gradient acceptability in HG.

4.1 Gradient acceptability in HG

One of the fundamental arguments for HG is that Harmony scores provide a means of modeling degrees of acceptability (Legendre et al. 1990a, 2006a). In the standard account (see also Keller 2000; 2006), degree of acceptability = $\mathcal{H}$. The abstract example in (29), shows how multiple violations of a single constraint leads to decreased $\mathcal{H}$, and hence decreased acceptability.

(29) Standard HG account: $\mathcal{H}$(Output-12) > $\mathcal{H}$(Output-22)

	Weight	1.5	1	$\mathcal{H}$
	Input-1	MARK	FAITH	
☞	Output-11		-1	-1
	Output-12	**-1**		**-1.5**

	Weight	1.5	1	$\mathcal{H}$
	Input-2	MARK	FAITH	
☞	Output-21		-2	-2
	Output-22	**-2**		**-3**

Boersma (2004) points out an apparently fatal flaw in this approach: because it does not relativize acceptability to the Input, in many cases ill-formed structures will receive higher scores than well-formed ones (see also Legendre et al. 2006b: 354 for a related conceptual critique of Legendre et al. 1990a, 2006a).

(30) A problem for the standard HG account: $\mathcal{H}$(Output-21) > $\mathcal{H}$(Output-12)

	Weight	3	2	1	$\mathcal{H}$
	Input-1	C1	C2	C3	
	Output-11	-2			-6
☞	**Output-12**		**-2**		**-4**

	Weight	3	2	1	$\mathcal{H}$
	Input-2	C1	C2	C3	
	Output-21		**-1**		**-2**
☞	Output-22			-1	-1

A 'real-life' example using Lombardi's (1999) analysis of final devoicing (see Boersma for syntactic examples):

(31) A problem for the standard HG account: $\mathcal{H}$([pad]) > $\mathcal{H}$([bam.bam])

Weight	3	2	1	$\mathcal{H}$
/bambam/	IDENT-ONS-VCE	*VOICE	ID-VCE	
[pam.pam]	-2		-2	-8
[pam.bam]	-1	-1	-1	-6
☞ **[bam.bam]**		**-2**		**-4**

Weight	3	2	1	$\mathcal{H}$
/pad/	IDENT-ONS-VCE	*VOICE	ID-VCE	
[pad]		**-1**		**-2**
☞ [pat]			-1	-1

To make gradient ill-formedness comparative and relative, I propose that Acceptability is the difference between the Harmony of a mapping and the most harmonic alternative:

(32) Acceptability(x) = $\mathcal{H}(x)$ - $\mathcal{H}(y)$

Where x is a candidate mapping, and y is the most harmonic mapping for the same Input, and where $x \neq y$

Using this metric, grammatical forms get positive scores, and ungrammatical ones get negative scores:

(33) Acceptability([bam.bam]) = $\mathcal{H}$([bam.bam]) - $\mathcal{H}$([pam.bam]) = **2**

Acceptability([pad]) = $\mathcal{H}$([pad]) - $\mathcal{H}$([pat]) = **-1**

Cumulative ill-formedness continues to be predicted under the revised HG Acceptability metric, as the following calculation for (29) shows:

(34) Acceptability(Output-12) = $\mathcal{H}$(Output-12) - $\mathcal{H}$(Output-11) = **-0.5**

Acceptability(Output-22) = $\mathcal{H}$(Output-22) - $\mathcal{H}$(Output-21) = **-1**

As far as I know, there is no current OT account of gradient acceptability that addresses this problem, and deals with both gradient ill-formedness and well-formedness.

4.2 Gradient acceptability in OT

Coetzee's (2004) account compares performance on the constraint hierarchy for Output candidates from different Inputs (see also Everett and Berent 1998). As such, it has the same weakness as the standard HG account:

(35) A problem for cross-tableau comparison

	IDENT-ONS-VCE	*VOICE	ID-VCE
/bambam/→ [bam.bam]		**	
/pad/→ [pad]		*	

OT accounts of gradient acceptability that are Input-relative involve considerable elaboration of the theory, and work only on one side of the ill-formed/well-formed divide:

(36) a. The account of gradient well-formedness in Pater (2005) involves the postulation of lexically specific faithfulness constraints for every lexical item, and requires the calculation of the outcome of every possible indexation to determine degree of well-formedness for a nonce word. As Hayes and Wilson (2006) point out, it does not extend to gradient ill-formedness.

b. The account of gradient ill-formedness in Boersma (2004) involves the incorporation of numerical values of constraints, and the calculation of the percentage of times each candidate wins relative to the other candidates. This account does not extend to gradient well-formedness, since by definition attested (non-variable) forms win 100% of the time.

5 Gradual HG learning, cumulativity, and gradience

One of the advantages of HG over OT is that it is compatible with the broad range of learning algorithms (and estimation and optimization procedures) that have been developed for linear models.

A particularly simple such algorithm is the Perceptron update rule of Rosenblatt (1958). Like the Gradual Learning Algorithm for stochastic OT (GLA; Boersma 1998; Boersma and Hayes 2001), and some implementations of the constraint demotion algorithm (Tesar and Smolensky 1998, 2000), Perceptron is on-line and error-driven. Given the learner's error and the correct observed mapping, the weight of each constraint is updated as follows:

(37) Add $n(v\text{E} - v\text{C})$ to the weight of each constraint

Where $0 < n < 1$, vE = number of violations incurred by the error, vC = number of violations incurred by the correct form

For comparison's sake, the GLA can be stated as:

(38) Add n to the weight of each constraint for which $v\text{E} > v\text{C}$, and subtract n from the weight of each constraint for which $v\text{E} < v\text{C}$

In the GLA, the degree of difference between violations on a constraint in the error and correct form is irrelevant (as it should be for OT evaluation).

The Perceptron coefficient adjustment procedure is similar to that used in:

(39) GLM regression analysis (see Keller 2006)

Stochastic gradient ascent for log-linear models of probability (see Jäger to appear)

Connectionist descendents of Perceptron (see papers in Smolensky and Legendre 2006)

The Perceptron model has a convergence/correctness proof, which is extended to part-of-speech tagging by Collins (2002). Collins' proof seems to further extend to an application of Perceptron with HG, and testing indicates that this grammar/learning model does correctly converge, even when the HG grammar has noisy evaluation, as in stochastic OT (Boersma and Pater in prep.). The GLA/stochastic OT combination is demonstrably non-convergent (Pater 2008).

5.1 Gradual HG learning and cumulativity

Gradual learning with HG predicts the emergence of cumulative constraint interaction, as we saw in Japanese loanwords. This can be modeled in Praat 4.6 (Boersma and Weenink 2007), in which the learner can be run with HG evaluation, and with the learning rule in (37).

The learning data were three representative words, distributed with equal frequency. At the initial state, the learner made errors on all three:

	ranking value	*disharmony*
***Vce–Gem**	100.000	100.000
OCP–Voice	100.000	100.000
Id–Voice	10.000	10.000

bobu	*Vce–Gem	OCP–Voice	Id–Voice	
bobu		*		100.000
☞ pobu			*	10.000

webbu	*Vce–Gem	OCP–Voice	Id–Voice	
webbu	*			100.000
☞ weppu			*	10.000

guddo	*Vce–Gem	OCP–Voice	Id–Voice	
guddo	*	*		200.000
☞ gutto			*	10.000

I ran the simulation with settings similar to those of (Boersma and Levelt 2000): plasticity was fixed at 0.1 throughout, which for the learning rule in (37) means $n = 0.1$.

The difference from the Boersma and Levelt simulation is HG evaluation, whose results can be seen in the state of the grammar after 1000 pieces of learning data:

	ranking value	*disharmony*
Id–Voice	80.004	79.017
OCP–Voice	49.368	49.422
***Vce–Gem**	48.251	44.425

bobu	Id–Voice	OCP–Voice	*Vce–Gem	
☞ bobu		*		–49.422
pobu	*			–79.017

webbu	Id–Voice	OCP–Voice	*Vce–Gem	
☞ webbu			*	–44.425
weppu	*			–79.017

guddo	Id–Voice	OCP–Voice	*Vce–Gem	
guddo		*	*	–93.848
☞ gutto	*			–79.017

Thus, the Japanese loanword pattern is a predicted stage of gradual HG learning. See Jesney and Tessier (in prep.) for results on modeling stages of child language acquisition.

5.2 Gradual learning and frequency-based gradience

A weighted constraint grammar that is learned gradually will reflect frequency differences in the relative weights of constraints. Even when these differences don't effect the outcome in I-O mappings, they can affect acceptability scores, both for the observed forms, and for ill-formed candidates (see relatedly Zuraw 2000).

A simple hypothetical case of a frequency skewing:

(40) In onset clusters, C2 is coronal 91% of the time

I submitted the following distribution of onset clusters to Praat with the standard settings for learning (except for HG evaluation/learning).

(41) Frequency of input data

klV 91%

kwV 9%

The markedness constraints (C2-Oral, C2-Lab, C2-Cor) began at 100, and the faithfulness constraint began at 10. The following shows the resulting grammar, and its evaluation of the attested forms, as well as two ill-formed onsets:

	ranking value	*disharmony*
C2–Oral	100.000	99.819
Faith	85.469	87.714
C2–Cor	71.975	70.653
C2–Lab	52.556	54.715

klV	C2–Oral	Faith	C2–Cor	C2–Lab	
☞ klV				*	–54.715
null		*			–87.714

kwV	C2–Oral	Faith	C2–Cor	C2–Lab	
☞ kwV			*		–70.653
null		*			–87.714

knV	C2–Oral	Faith	C2–Cor	C2–Lab	
knV	*			*	–154.534
☞ null		*			–87.714

kmV	C2–Oral	Faith	C2–Cor	C2–Lab	
kmV	*		*		–170.472
☞ null		*			–87.714

Both the relative Acceptability of the attested forms, and of the ill-formed clusters, is affected by the weighting C2-Lab > C2-Cor. This highlights an important difference between this phonological approach to gradient phonotactics, and a theory that relied only on raw segmental probability (see relatedlyMoreton 2002; Albright 2006; Hayes and Wilson 2006):

(42) Generalization based on phonological constraints leads to predicted asymmetries in structures of equal frequency, based on distributional differences elsewhere in the language, and on universals that are encoded in the constraint set

6 Conclusions

To return to the main point of this talk, we can look at one of Ohala and Ohala's stimulus pairs in terms of the HG account of gradient acceptability

(43) Cumulative ill-formedness in HG

Weight	100	10	$\mathcal{H}$
/xlox/	*x	Ident-Cont	
[xlox]	-2		-200
☞ [klok]		-2	-20

Acceptability(xlox) = -180

Weight	100	10	$\mathcal{H}$
/xlit/	*x	Ident-Cont	
[xlit]	-1		-100
☞ [klit]		-1	-10

Acceptability(xlit) = -90

[xlox] would be rated worse than [xlit]: this fits Ohala and Ohala's results.

The typological prediction is that a language that hardens a velar fricative in the presence of another one in the same word is impossible, since it would require contradictory weighting conditions:

(44) No cumulative ill-formedness in HG

/xlox/ → [klok]: w(*x) > w(Ident-Cont)

/xlit/ → [xlit]: w(Ident-Cont) > w(*x)

This seems to fit with what we find typologically.

Much remains to be done in terms of further developing and evaluating HG as a theory of typology and of gradient acceptability, as well as better understanding the nature of cumulativity in different domains.

References

Albright, Adam. 2006. Gradient phonotactic effects: lexical? grammatical? both? neither? Albuquerque. LSA 2006.

Alderete, John. 1997. Dissimilation as local conjunction. In *Proceedings of the north east linguistic society 27*, ed. Kiyomi Kusumoto and Kiyomi Kusumoto, 17–32. Amherst, Mass.: GLSA Publications.

Berent, Iris, Daniel Everett, and J. Shimron. 2001. Do phonological representations specify formal variables? Evidence from the Obligatory Contour Principle. *Cognitive Psychology* .

Berent, Iris, and J. Shimron. 1997. The representation of Hebrew words: Evidence from the Obligatory Contour Principle. *Cognition* 64:39–72.

Berent, Iris, Donca Steriade, Tracy Lennertz, and Vered Vaknin. 2006. What we know about what we have never heard: Evidence from perceptual illusions. *Cognition* .

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

Boersma, Paul. 2004. A stochastic ot account of paralinguistic tasks such as grammaticality and prototypicality judgments Ms, University of Amsterdam.

Boersma, Paul, and Bruce Hayes. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32:45–86. Available on Rutgers Optimality Archive, http://ruccs.rutgers.edu/roa.html.

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 and Eve V. Clark. Stanford, CA: CSLI Publications. Available on Rutgers Optimality Archive, http://ruccs.rutgers.edu/roa.html.

Boersma, Paul, and David Weenink. 2007. Praat: Doing phonetics by computer, version 4.6. URL `http://www.praat.org/`, Developed at the Institute of Phonetic Sciences, University of Amsterdam.

Chomsky, Noam, and Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper & Row.

Coetzee, Andries. 2004. What it means to be a loser: Non-optimal candidates in Optimality Theory. Ph. D dissertation, UMass Amherst.

Collins, Michael. 2002. Discriminative training methods for hidden markov models: Theory and experiments with perceptron algorithms. *Proceedings of the Conference on Empirical Methods in Natural Language Processing* .

Everett, Daniel, and Iris Berent. 1998. The comparative optimality of hebrew roots: An experimental approach to violable identity constraints. Ms, University of Pittsburgh and Florida Atlantic University.

Frisch, Stefan, and Bushra Zawaydeh. 2001. The psychological reality of ocp-place in arabic. *Language* 77:91–206.

Hay, Jennifer, Janet Pierrehumbert, and Mary Beckman. 2004. Speech perception, well-formedness, and the statistics of the lexicon. In *Papers in laboratory phonology vi*, 58–74. Cambridge, UK: Cambridge University Press.

Hayes, Bruce, and Colin Wilson. 2006. A maximum entropy model of phonotactics and phonotactic learning. Ms., UCLA.

Ito, Junko, and Armin Mester. 1986. The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility. *Linguistic Inquiry* 17:49–73.

Ito, Junko, and Armin Mester. 1998. German coda conditions: Sympathy meets freedom of the input. Tuebingen, Germany. The Syllable: Typology and Theory.

Ito, Junko, and Armin Mester. 2003. *Japanese morphophonemics: Markedness and word structure*. Cambridge, Massachusetts: MIT Press Linguistic Inquiry Monograph Series 41.

Jäger, Gerhard. to appear. Maximum entropy models and Stochastic Optimality Theory. In *Architectures, rules, and preferences: A festschrift for Joan Bresnan*, ed. Jane Grimshaw, Joan Maling, Chris Manning, Jane Simpson, and Annie Zaenen. Stanford, CA: CSLI.

Kawahara, Shigeto. 2006. A faithfulness ranking projected from a perceptibility scale: The case of [+voice] in japanese. *Language* 82:536–574.

Keller, Frank. 2000. Gradience in grammar: Experimental and computational aspects of degrees of grammaticality. Doctoral Dissertation, University of Edinburgh.

Keller, Frank. 2006. Linear optimality theory as a model of gradience in grammar. In *Gradience in grammar: Generative perspectives*, ed. Gisbert Fanselow, Caroline Féry, Ralph Vogel, and Matthias Schlesewsky. Oxford: Oxford University Press.

Legendre, Geraldine, Yoshiro Miyata, and Paul Smolensky. 1990a. Harmonic Grammar – A formal multi-level connectionist theory of linguistic well-formedness: An Application. In *Proceedings of the twelfth annual conference of the cognitive science society*, 884–891. Mahwah, NJ: Lawrence Erlbaum Associates.

Legendre, Géraldine, Yoshiro Miyata, and Paul Smolensky. 1990b. 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, Yoshiro Miyata, and Paul Smolensky. 1990c. Harmonic Grammar – a formal multi-level connectionist theory of linguistic wellformedness: Theoretical foundations. In *Proceedings of the twelfth annual conference of the Cognitive Science Society*, 388–395. Cambridge, MA: Lawrence Erlbaum.

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

Legendre, Géraldine, Antonella Sorace, and Paul Smolensky. 2006b. The Optimality Theory–Harmonic Grammar connection. In Smolensky and Legendre (2006), 339–402.

Lombardi, Linda. 1999. Positional faithfulness and voicing assimilation in Optimality Theory. *Natural Language and Linguistic Theory* 17:267–302.

Lubowicz, Ania. 2005. Locality of Conjunction. In *Proceedings of WCCFL 24*. Somerville, MA: Cascadilla.

McCarthy, John J. 1999. Sympathy and phonological opacity. *Phonology* 16:331–399.

McCarthy, John J. 2003. Comparative Markedness. *Theoretical Linguistics* 29:1–51.

Moreton, Elliott. 2002. Structural constraints in the perception of english stop-sonorant clusters. *Cognition* 84:55–71.

Moreton, Elliott, Gary Feng, and Jennifer L. Smith. 2005. Syllabification, sonority, and perception: New evidence from a language game. In *Proceedings of the 41st Regional Meeting of the Chicago Linguistic Society*. Http://www.unc.edu/ moreton/Papers/MoretonFengSmithCLS2005.pdf.

Moreton, Elliott, and Paul Smolensky. 2002. Typological consequences of local constraint conjunction. In *The proceedings of the 21st west coast conference in formal linguistics*, ed. L. Mikkelsen and C. Potts. Somerville, MA: Cascadilla Press.

Nishimura, Kohei. 2003. Lyman's law in loanwords Ms, University of Tokyo.

Ohala, John J., and Manjari Ohala. 1986. Testing hypotheses regarding the psychological manifestation of morpheme structure constraints. In *Experimental phonology*, ed. John J. Ohala and Jeri J. Jaeger, 239–252. Orlando: Academic Press.

Pater, Joe. 2005. Learning a stratified grammar. In *Proceedings of the 29th boston university conference on language development*, ed. Alejna Brugos, Manuella R. Clark-Cotton, and Seungwan Ha, 482–492. Somerville, MA: Cascadilla.

Pater, Joe. 2006. Additive optimization and phonological typology. URL `http://people.umass.edu/pater/AO+typology.pdf`, Handout from a talk presented at the MIT/UMass Phonology Workshop, February 11.

Pater, Joe. 2007. The Power of Constraint Weighting Ms, University of Massachusetts, Amherst.

Pater, Joe. 2008. Gradual learning and convergence. *Linguistic Inquiry* To appear.

Pater, Joe, and Andries Coetzee. 2005. Lexically specific constraints: Gradience, learnability, and perception. In *Proceedings of the 3rd seoul international conference on phonology*, 85–119. Seoul: The Phonology-Morphology Circle of Korea.

Pertz, D. L., and T. G. Bever. 1975. Sensitivity to phonological universals in children and adolescents. *Language* 51:149?162.

Pierrehumbert, Janet. 2001. Stochastic phonology. *GLOT* 5:1–13.

Prince, Alan, and Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Malden, Mass., and Oxford, UK: Blackwell. Published version of.

Rose, Sharon, and Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language* 80:475–532.

Rosenblatt, Frank. 1958. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review* 65:386–408.

Scholes, Robert. 1966. *Phonotactic grammaticality*. The Hague: Mouton.

Smolensky, Paul. 2006. Optimality in phonology II: Harmonic completeness, local constraint conjunction, and feature domain markedness. In Smolensky and Legendre (2006), 27–160.

Smolensky, Paul, Lisa Davidson, and Peter Jusczyk. 2003. The Initial and Final States: Theoretical Implications and Experimental Explorations of Richness of the Base. In *Fixing priorities: constraints in phonological acquisition*, ed. Rene Kager, Joe Pater, and Wim Zonneveld. Cambridge: Cambridge University Press.

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

Suzuki, Keiichiro. 1998. A Typological Investigation of Dissimilation. Doctoral Dissertation, University of Arizona, Tucson, AZ.

Tesar, Bruce, and Paul Smolensky. 1998. Learnability in Optimality Theory. *Linguistic Inquiry* 29:229–268.

Tesar, Bruce, and Paul Smolensky. 2000. *Learnability in Optimality Theory*. Cambridge, Mass.: MIT Press.

Zuraw, Kie. 2000. Patterned exceptions in phonology. Doctoral Dissertation, UCLA.