In *Proceedings of the 3rd Seoul International Conference on Phonology*. Seoul: The Phonology-Morphology Circle of Korea. 85-119.

Lexically Specific Constraints: Gradience, Learnability, and Perception*

Joe Pater and Andries Coetzee
University of Massachusetts, Amherst and University of Michigan

1. Introduction

Lexically specific constraints are indexed versions of constraints that apply only when a morpheme that bears that index is evaluated by the grammar. They have been used in Optimality Theory to deal with exceptions (e.g. Hammond 1995, Kraska-Szlenk 1997, Pater 2000), and have also been applied to the lexical strata of Japanese and other languages (Fukuzawa 1999, Itô and Mester 1999, 2001, Gelbart 2005). In this paper, we propose a further application of lexically specific constraints: to the analysis of gradient phonotactics (cf. Frisch, Pierrehumbert and Broe 2004). Markedness constraints are ranked according to the degree to which they are obeyed across the words of the language, with lexically specific constraints interspersed between them. We then show that rankings of this type can be learned with a relatively minor elaboration of the Biased Constraint Demotion Algorithm (Prince and Tesar 2004). Finally, we provide experimental evidence from lexical decision tasks and acceptability judgments that language users are aware of such lexical patterns.

2. Lexically specific constraints and gradient phonotactics

2 .1 Exceptions and lexically specific constraints

In Chomsky and Halle (1968) and subsequent research in generative phonology, exceptions have been dealt with in two ways. In *structural* analyses, exceptional lexical items are specified with phonological structure that is not present on regular items. For example, in English, nouns with light penultimate syllables are generally stressed on the antepenult (e.g. *Cánada*). Exceptions like *banána* could have an underlying accent on the penult that blocks the application of the rule or constraint that would usually place stress on the penult. In *diacritic* analyses, exceptional lexical items are marked as subject to a rule or constraint that does not apply to

* Special thanks to René van den Berg for sharing an electronic version of his Muna dictionary with us (Berg and Sidu 1996), and for his comments on parts of this research. Portions of this work were presented at the University of Victoria, New York University, the Boston University Conference on Language Development, the Manchester Phonology Meeting and at the University of Edinburgh. We'd like to thank the participants for helpful discussion, and especially the following people at those events and elsewhere: Adam Albright, Lisa Davidson, Kathryn Flack, Stefan Frisch, Diamandis Gafos, Ben Gelbart, Greg Guy, Mike Hammond, Bruce Hayes, Shigeto Kawahara, John Kingston, John McCarthy, Elliot Moreton, Mits Ota, Steve Parker, Lisa Selkirk, Bruce Tesar, Anne-Michelle Tessier and Su Urbanczyk.

regular items (or vice versa). An application of this approach would be to mark *banána* as an exception to final syllable extrametricality, so that the final two syllables are placed into a foot, unlike regular items, in which the final syllable would be skipped in footing.

In Optimality Theory (OT; Prince and Smolensky 1993/2004), exceptions have continued to be dealt with in these two ways (see Pater 2004 for references). The structural analysis involves positing a faithfulness constraint that dominates the relevant markedness constraint. In the *banána* case, preserving the lexical accent on the penultimate syllable will entail violating at least one markedness constraint that would prefer antepenultimate stress for the regular *Canada*-type words. Assuming that *banána* is footed with a final trochee, Prince and Smolensky's replacement for extrametricality, NONFINALITY, will be violated. It must be dominated by a constraint demanding preservation of lexical stress, which I will call STRESS-FAITH (Pater 2000). In (1) the tableaux for regular *Cánada* and exceptional *banána* are provided (parentheses indicate foot boundaries):

(1) **Exceptionality as faithfulness**

banána	STRESS-FAITH	NONFINALITY
☞ ba(nána)		*
(bána)na	* !	

Canada	STRESS-FAITH	NONFINALITY
☞ (Cána)da		
Ca(náda)		*!

When a lexical accent is present, as in *banana*, preservation of that accent overrides the requirement that the final syllable not be incorporated into the head foot. When there is no underlying accent, STRESS-FAITH is irrelevant, and NONFINALITY can be satisfied, as in *Canada*.

A diacritic analysis of exceptionality in OT involves a lexically specific markedness or faithfulness constraint that applies only to lexical items that are indexed for its application.[1] In this case, we could appeal to a lexically specific version of ALIGN-RIGHT, the constraint that demands that a foot be in final position, which conflicts with NONFINALITY. The lexically specific ALIGN-RIGHT dominates NONFINALITY; the general one is ranked beneath. The result is shown in (2), in which *banana* and the lexically specific ALIGN-RIGHT are indexed with the

[1] A related diacritic approach is for lexical items to select rankings of constraints, or "co-phonologies" (see e.g. McCarthy and Prince 1993, Nouveau 1994, Itô and Mester 1995, Inkelas, Orgun and Zoll 1997, Anttila 2002, Inkelas and Zoll 2003, Caballero 2005, Zamma 2005). When both markedness and faithfulness constraints are allowed to be indexed (Pater 2000, Gelbart 2005, cf. Fukuzawa 1999, Itô and Mester 1999, 2001), these approaches are nearly equivalent, though see Pater (2004, 2005) on the advantages of indexed constraints.

diacritic ‘L’. *Canada* is not marked with the diacritic, so ALIGN-RIGHT-L does not apply.

(2) **Exceptionality as lexically indexed markedness**

$banana_L$	ALIGN-RIGHT-L	NONFINALITY	ALIGN-RIGHT
☞ ba(nána)		*	
(bána)na	* !		*

Canada	ALIGN-RIGHT-L	NONFINALITY	ALIGN-RIGHT
☞ (Cána)da		* !	
Ca(náda)			*

If all else were equal, the structural account of exceptionality would be preferable. It makes use only of independently motivated phonological structure, and does not require the proliferation of cloned constraints.

One argument for a diacritic account of exceptions, which we will be building on in this paper, is that in many cases a purely structural analysis does not distinguish between a regular and an exceptional pattern. Languages often restrict marked structures to a small set of exceptional words, most commonly recent borrowings (see e.g. Itô and Mester 1999), but sometimes also native words (as in Latvian; Gelbart 2005). Without diacritics, one cannot usually distinguish between such a marginal structure, and one that is fully general. To take an extreme hypothetical example, the grammar of a language with only one word with a coda would have the same ranking as one with no restriction on codas: FAITH >> NOCODA, where ‘FAITH’ stands for the set of faithfulness constraints that conflict with NOCODA. Without lexically specific constraints, or some other grammatical mechanism for exceptionality, even a single word with a coda would force this ranking for the whole language.

With lexically specific faithfulness, however, the general absence of codas from the “one coda” language can be expressed with the ranking of NOCODA over a faithfulness constraint like MAX, which produces consonant deletion. A lexically specific MAX-L would then protect the coda in the exceptional word. A grammar and lexicon for this hypothetical language are shown in (3).

(3) Grammar: MAX-L >> NOCODA >> MAX
Lexicon: /ma/ /sa/ /fi/ /no/ /ka/ /la/ /se/ /te/ /kit_L/

The grammatical encoding of the exceptionality of marginal structures is supported by speakers’ awareness of the difference between regular and marginal patterns (see section 4), as well as by the tendency of exceptional patterns to be regularized.

Importantly, the addition of lexically specific constraints to the theory does not mean that “anything goes”. The constraints are not arbitrary: lexically specific constraints are indexed versions of general constraints (cf. Hammond 1995: 15), which as we argue in section 3, are constructed in the course of learning (see also Pater 2004). Furthermore, the ranking of these constraints in the grammar captures the range of possible exceptions. For example, in our hypothetical language in (3), the constraint *COMPLEX ("no consonant clusters") would dominate MAX-L, since there is no evidence to contradict the preferred Markedness >> Faithfulness ranking (Smolensky 1996, Hayes 2004, Prince and Tesar 2004):

(4) Grammar: *COMPLEX >> MAX-L >> NOCODA >> MAX
Lexicon: /ma/ /sa/ /fi/ /no/ /ka/ /la/ /se/ /te/ /kit_L/

If by Richness of the Base (Prince and Smolensky 1993/2004) this grammar were supplied with a word with consonant cluster, it would be reduced, even if the word were indexed to MAX-L (e.g. /bla_L/ -> [ba]).

Itô and Mester (1999, 2001) extend this approach to account for the implicational patterns of exceptionality found amongst the sets of words in Japanese and other languages (see also Fukuzawa 1999), and argue that these implications cannot be captured in a structural account. In this paper, we apply lexically specific constraints to account for graded exceptionality, a phenomenon that is also out of the reach of a structural account, since it is a more extreme case of the problem raised by the "one coda" language. The grammar in (4) expresses three degrees of acceptability: acceptable (CV words), unacceptable (CCV words), and exceptional (CVC words). However, it is possible to have exceptions that differ in the degree of attestedness, and hence acceptability. In the next section, we discuss a well-known case of this: restrictions on homorganic consonants in the Arabic verbal root system.

2.2 Arabic and the problem of lexical gradience

In Arabic, there is a restriction against homorganic consonants in adjacent positions within the verbal root (Greenberg 1950, McCarthy 1988, 1994, Pierrehumbert 1993, Frisch *et al.* 2004). Given Arabic's root-and-pattern morphological system, these consonants are often separated by vowels supplied by other morphemes, so they are not necessarily adjacent in the phonological string. There are a number of exceptions to this restriction, but the distribution of these exceptions is not random. Frisch *et al.* (2004) examined the degree to which particular pairs of consonants are underrepresented in a list of 2,674 roots taken from a dictionary of standard Arabic (Cowan 1979). For each pair, they calculated an Observed/Expected ratio, that is, the number of observed words of a particular type divided by the number that would be expected if the members of the pair on consonants co-occurred freely. The lower the O/E value, the more underrepresented a pair is. Frisch *et al.*'s results, collapsed over place of articulation, are shown in Table 1. It should be noted that they omitted pairs with identical consonants, which behave differently; we have also omitted the pharyngeals, since these are irrelevant to a comparison with Muna, the language we discuss in section section 2.3.

	Labial b f m	**Dorsal** k g q	**Coronal Sonorant** l r n	**Coronal Fricative** θ ð s z sˤ zˤ ʃ	**Coronal Plosive** t d tˤ dˤ
Labial	0.00				
Dorsal	1.15	0.02			
Coronal Sonorant	1.18	1.48	0.06		
Coronal Fricative	1.31	1.16	1.21	0.04	
Coronal Plosive	1.37	0.80	1.23	0.52	0.14

Table 1 O/E for "adjacent" consonants in Arabic verbal roots

O/E values for pairs of homorganic consonants in this table are generally near zero. However, within the coronals, there are several degrees of representation. Pairs of coronals that are both plosives, both fricatives, or both sonorants are highly underrepresented. Sonorants co-occur with plosives and fricatives at a rate slightly higher than expected. Fricative-plosive pairs co-occur at an intermediate rate.

McCarthy (1988) accounts for the split between sonorants and obstruents by specifying that the constraint against homorganic segments, the OCP for consonantal place, applies only within the subclasses of coronals defined by the manner feature [+/-sonorant]. In Dresher (1989), Selkirk (1991), and Padgett (1995), the general OCP constraint is elaborated into a set of more specific constraints that are violated only by segments that are identical in particular ways (see especially Padgett 1995 on this aspect of the proposal). The relativized OCP constraints that would correspond to McCarthy's (1988) groupings appear in (5).[2] In McCarthy's analysis, the consonants in question are assumed to be adjacent at a derivational level in which intervening vowels are absent.

[2] McCarthy (1988) also divides the dorsals by sonority, classifying /w, y/ as the dorsal sonorants, though he states that their lack of co-occurrence may have other explanations.

(5) OCP-LAB Adjacent labials are prohibited
OCP-DOR Adjacent dorsals are prohibited
OCP-PHAR Adjacent pharyngeals are prohibited
OCP-COR[αSON] Adjacent coronals agreeing in [+/-sonorant] are prohibited

These constraints are categorical: the grammar bans pairs of consonants that fall under their scope, and permits ones that do not. With OCP-COR relativized to [+/-sonorant], obstruent/sonorant pairs of coronals are permitted, while pairs of obstruents and pairs of sonorants are banned.

The difficulty is in creating a grammar compatible with the intermediate degree of representation of the fricative/stop pairs; given the constraints in (5), they should be ruled out. If the OCP-COR for obstruents is further specified to apply only between segments that have the same value for [+/-continuant], as in Padgett (1995), then there is no restriction on stop/fricative pairs. Neither of these seems correct. McCarthy (1994) states that continuancy restriction on the OCP-COR is "not absolute", though this is not formalized.

The ranked and violable constraints of Optimality Theory offer a way out of this dilemma, but not in their standard interpretation. In (6), a further elaborated set of OCP-COR constraints is presented. We continue to use the term 'adjacent', but under the standard OT assumption that markedness constraints apply at the surface, adjacency must be defined so as to ignore intervening vowels (see Suzuki 1998, Rose 2000):

(6) OCP-COR No adjacent coronals
OCP-COR[+SON] No adjacent coronal sonorants
OCP-COR[-SON] No adjacent coronal obstruents
OCP-COR[-SON][αCONT] No adjacent coronal obstruents agreeing in continuancy

The ranking of these constraints that is appropriate for Arabic is as in (7):

(7) OCP-COR[-SON][αCONT], OCP-COR[+SON] >> OCP-COR[-SON] >> OCP-COR

Pairs of segments that meet the definition of the constraints in the top-most stratum are highly underrepresented, with O/E values approaching 0. Pairs of segments that fall outside the scope of those constraints, but that do meet the definition of the constraint in the middle stratum have an O/E value of 0.52. Pairs of segments that meet only the definition of the lowest ranked constraint have an O/E value of above 1.

The problem is that neither standard OT, nor the versions that have been proposed to handle variation, allow the ranking in (7) to express an Arabic speaker's knowledge of the observed lexical gradience. For a markedness constraint to be violated, a conflicting constraint must outrank it. Because we are not dealing with alternations, we will use the undifferentiated faithfulness constraint in (8):

(8) FAITH The Input representation and the Output representation are identical

FAITH must be ranked above OCP-COR, so that obstruent-sonorant pairs surface intact. It must be ranked beneath the topmost stratum, so that sonorant pairs as well as obstruent pairs agreeing in continuancy will be altered to fit the demands of the markedness constraints (we temporarily abstract from the small set of exceptions). The issue is the ranking of FAITH with respect to OCP-COR[-SON]. Placed below it, stop-fricative pairs are ruled out. Placed above it, they are perfectly well-formed. We again face the quandary that neither of these seems correct.

If the ranking between these constraints is allowed to vary each time the grammar is employed, as in models like those of Anttila (1997) and Boersma (1998), then a word with a stop-fricative pair will vary between a faithful output, and one that is altered (for example by deleting, or changing the place, of one of the segments). But lexical items in Arabic that have stop-fricative pairs are not reported to show variation. Hammond's (2004) approach to lexically gradient acceptability does not distinguish between variation and exceptionality, and without further modification also leaves this issue unresolved. Zuraw (2000) does extend Boersma's (1998) theory to patterned exceptionality by having a USE-LISTED constraint block variation, but this account only covers cases involving alternation.

In a theory with lexically specific constraints, we can resolve this dilemma by having a specific version of FAITH that is indexed to words that contain stop-fricative pairs. This constraint FAITH-L ranks above OCP-COR[-SON], thus protecting them from its demands:

(9) OCP-COR[-SON][αCONT], OCP-COR[+SON] >> FAITH-L >> OCP-COR[-SON]
>> FAITH >> OCP-COR

The next step is to provide a means by which a speaker could compute the relative grammaticality of forms from the ranking in (9). So far, one might simply say that a speaker knows that forms with stop-fricative pairs must be lexically marked as exceptions, and that this gives them an intermediate status between forms that are ruled out completely (for which the relevant markedness constraints dominate FAITH-L), and those that are perfectly acceptable (for which the relevant markedness constraints are dominated by FAITH).

However, this would not be sufficient to deal with further degrees of gradience. In Arabic, forms that violate OCP-COR[-SON][αCONT] and OCP-COR[+SON] are very rare, but are attested. This means that faithfulness must rank above these constraints:

(10) FAITH-L2 >> OCP-COR[-SON][αCONT], OCP-COR[+SON] >> FAITH-L1 >> OCP-COR[-SON] >> FAITH >> OCP-COR

In the hierarchy in (10), another lexically specific faithfulness constraint, FAITH-L2, has been installed to allow for the exceptional forms with pairs of coronal fricatives, coronal stops, and coronal sonorants. Pairs of labials, on the other hand, are completely absent from the lexicon of Arabic roots. To rule these out, OCP-LAB must dominate faithfulness:

(11) OCP-LAB >> FAITH-L2 >> OCP-COR[-SON][αCONT], OCP-COR[+SON] >> FAITH-L1 >> OCP-COR[-SON] >> FAITH >> OCP-COR

We now have four grades of acceptability: ungrammatical (pairs of labials), marginally acceptable (e.g. pairs of coronal obstruent stops), moderately acceptable (pairs of coronal obstruents disagreeing in continuancy) and acceptable (pairs of coronals disagreeing in sonorancy). Ungrammaticality is expressed by the fact that a word with a pair of labials will never surface intact, no matter which faithfulness constraint it is indexed to. Perfect acceptability is expressed by the fact that a word with an obstruent-sonorant pair of coronals will always surface faithfully.

One way to distinguish intermediate grades is by submitting a word to the grammar with each lexical indexation. The more often it surfaces faithfully, the more acceptable it is (cf. Anttila's 1997 approach to variation). Given a word that contains a pair of coronal obstruent stops, indexing it to FAITH-L2 will allow it to surface faithfully, while indexing it to FAITH-L1, or leaving it unindexed, will force it to be altered to satisfy OCP-COR[-SON][αCONT]. That is, in 1/3 of the possible indexations, a pair of coronal obstruent stops will surface faithfully (12b). On the other hand, a pair of coronal obstruents disagreeing in continuancy will surface faithfully with 2/3 of the indexations: it will be altered only if it is left unindexed (12c). A pair of labials will never surface faithfully (12a), while a pair of coronals disagreeing in sonorancy will survive intact regardless of indexation (12d). The ordering that thus emerges is shown in (12e).[3]

(12)	a.	*p-m	*p-m_{L1}	*p-m_{L2}	0/3
	b.	*t-d	*t-d_{L1}	✓t-d_{L2}	1/3
	c.	*t-s	✓t-s_{L1}	✓t-s_{L2}	2/3

[3] This approach yields relative well-formedness of different structures depending on their lexical frequency, but does not directly mirror lexical frequency. For example, the "one coda" language discussed in 2.1, and a language with more exceptional coda-bearing words (say, three) could have the same grammars, and the same ratings for well-formedness of codas. Whether they do or not would depend on other patterns of exceptionality. For example, if these languages both had two words with a cluster, then the grammar for the one coda language would be MAX-L1 >> NOCODA >> MAX-L2 >> *COMPLEX >> MAX, whereas the three coda language would have MAX-L1 >> *COMPLEX >> MAX-L2 >> NOCODA >> MAX. Thanks to Mike Hammond, Bruce Hayes, and Mits Ota for raising this issue.

d. ✓t-n ✓t-n $_{L1}$ ✓t-n $_{L2}$ 3/3
e. t-n > t-s > t-d > p-m

Another possibility is to compute relative well-formedness by comparing tableaux for different forms (as suggested in Everett and Berent 1998; see Coetzee 2004 for a formally explicit proposal). A pair of coronal stops violates a higher ranked constraint (OCP-COR[-SON][αCONT]) than a pair of coronal obstruents disagreeing in continuancy (OCP-COR[-SON]). In Coetzee's rank-ordering model of Optimality Theory, the grammar can use this information to place forms into a harmonic ordering. Both of these approaches make use of the ranked and violable constraints of standard Optimality Theory to generate gradient well-formedness. They extend the theory in different ways, but neither is inconsistent with its use for categorical patterns.

2.3 Muna and the similarity metric

Frisch *et al.* (2004) provide an analysis of Arabic that predicts the relative rates of co-occurrence of pairs of homorganic consonants using a similarity metric. Similarity is calculated for homorganic pairs by dividing the number of natural classes that both segments belong to by the total number of natural classes that each segment belongs to. They argue that this analysis has three advantages over an OT account (p. 219):

(13) i. It allows for gradient well-formedness
ii. It has greater predictiveness: regardless of the feature combinations at issue, pairs that are equally similar should be equally underrepresented
iii. It also predicts an effect of inventory size on similarity

As for the first advantage, we demonstrated in the previous section that a version of OT that incorporates lexically specific constraints allows for a grammar that reflects lexically gradient patterns. This version of OT also accounts for categorical constraint interaction, which Frisch *et al.*'s theory is silent on (Frisch *et al.* 2004: 221-222). However, the other differences between the similarity account and the OT account continue to apply. To help to illustrate the second one, in (14) we repeat the partial constraint hierarchy that we have postulated for Arabic. Given factorial typology, it is predicted that these constraints like these could be ranked in other ways. In particular, there is no reason that sonorancy, rather than some other feature, should be the main determinant of the strength of the OCP-COR restriction. The similarity metric, on the other hand, predicts that pairs of segments that are equivalent in similarity should be equally underrepresented, which would mean that particular OCP constraints should always be grouped together.[4]

(14) OCP-Lab >> Faith-L2 >> OCP-Cor[-son][αcont], OCP-Cor[+son] >> Faith-L1 >> OCP-Cor[-son] >> Faith >> OCP-Cor

The third difference that Frisch *et al.* (2004) point to is the fact that the similarity metric derives the fact that the co-occurrence restriction is weakened in coronals from the large size of the Arabic coronal inventory. The OT account does not relate the weakness of the OCP effect in coronals to inventory size.

In this section, we argue that these differences may in fact favor the OT analysis, given the patterns of consonantal place co-occurrence in Muna, an Austronesian language spoken on an island near Sulawesi. In his grammar of Muna, van den Berg (1989) examined the consonantal co-occurrence patterns in a set of 1100 CVCV roots, and noted that homorganic stops that differ in voicing do not co-occur, nor do nasals and homorganic obstruents. These restrictions are suggestive of a broader generalization: a ban on non-identical homorganic segments, as in Semitic (Greenberg 1950) and Javanese (Uhlenbeck 1949). Here we present some of the results of a study of the patterns in the 5854 CVCV and CVCVCV roots in an electronic version of Berg's dictionary; for further details see Coetzee and Pater (2005). The consonantal inventory of Muna is presented in (15).

(15)	*labial*	*coronal*	*velar*	*uvular*	*glottal*
voiceless	p	t	k		
voiced	b	d	g		
implosive	ɓ	ɗ			
nasal	m	n	ŋ		
-vce, prenas	^{m}p	^{n}t	^{ŋ}k		
		^{n}s			
+vce, prenas	^{m}b	^{n}d	^{ŋ}g		
-vce, fric	f	s			h
+vce, fric				ʁ	
trill		r			
lateral		l			
glide	w				

Like many Austronesian languages, Muna has an absolute ban on multiple prenasalized stops. The following table presents the overall results for non-identical

[4] The similarity account does not in fact predict the overwhelming effect of sonorancy either (Frisch *et al.* 2004: 204), but it does make predictions about relative rates co-occurrence that the relativized OCP analysis does not.

consonants in "adjacent" positions (C_1/C_2 and C_2/C_3 in $C_1VC_2V(C_3V)$), with pairs of prenasalized stops omitted (the uvular fricative is included with the dorsals; the glottal fricative is omitted).[5]

	Labial			Dorsal			Coronal		
	Obs	Exp	**O/E**	Obs	Exp	**O/E**	Obs	Exp	**O/E**
Labial	132	442	**0.30**						
Dorsal	875	771	**1.14**	29	164	**0.18**			
Coronal	2741	2180	**1.26**	1766	1523	**1.16**	1338	1686	**0.79**

Table 2 O/E for "adjacent" non-identical consonants in Muna roots

As in Arabic, the homorganic consonants are underrepresented at all places of articulation, and like Arabic, the coronal pairs are less underrepresented than those of the other places are. However, there are also important differences between Muna and Arabic. Table 3 presents the O/E values for the Muna coronals that can be compared with Arabic (it omits the implosive stop, and the prenasalized stops and fricatives).

	t	**d**	**s**	**l**	**r**	**n**
t	1.02					
d	0.60	2.54				
s	0.37	0.55	1.21			
l	0.78	0.79	1.13	0.89		
r	0.88	0.84	1.08	0.19	0.45	
n	0.70	0.25	1.17	0.32	0.56	2.94

Table 3 O/E values for Muna coronals

There is not a clear obstruent/sonorant split as in Arabic. In Muna, the voiced stop co-occurs more freely with the nasal than with the voiceless stop; this is true of the other places of articulation as well. This is unexpected under the similarity metric. Table 4 presents the O/E values and the average values for similarity calculated according to Frisch *et al.*'s (2004) metric (see Coetzee and Pater 2005 for details). Similarity should correlate negatively with O/E, but here they are correlated positively.

[5] Identical consonants co-occur freely in all positions in Muna. In addition, pairs of segments that are identical except that one of the segments is prenasalized also seem to be exempt from the co-occurrrence restriction; these have been omitted from all of the tables here. See Coetzee and Pater (2005) for further discussion.

	Voiced Stops+ Nasals	Voiced Stops+ Voiceless Stops
O/E	0.15	0.32
Similarity	0.22	0.33

Table 4 O/E and average similarity values

The low co-occurrence rate of the nasals and voiced stops seems to be due to their shared voicing; nasals and voiceless stops are not nearly as underrepresented. Along with voicing, continuancy and sonorancy also play a role in determining co-occurrence rates amongst the Muna coronals. These factors have all been identified as playing a role in the Arabic co-occurrence constraints (see Frisch *et al.* 2004 on voicing). However, the relative weighting of these factors in Muna and Arabic is different.

In (16), we present a constraint ranking that reflects the relative co-occurrence rates in Table 3. Beside each markedness constraint are the consonant pairs that it crucially targets (i.e. those that do not violate a higher ranked constraint). For the purposes of this analysis, we have grouped together pairs with O/E values from 0 to .30, which are targeted by the constraint in the highest markedness stratum, those with values between .31 and .69, which are targeted by the constraint in the intermediate stratum, and those with values of .70 and above, which are targeted only by the general OCP-COR. These groupings are necessarily arbitrary, since we have no information on acceptability judgments by Muna speakers. Further distinctions could be made by including further constraints, but the simple hierarchy in (16) is useful for a comparison with the Arabic hierarchy in (14).

(16) **OCP-CORONAL sub-hierarchy for Muna**

FAITH-L2 >> OCP-COR[αCONT][αVOICE](nd, lr) >> FAITH-L1 >>
OCP-COR[αSON](sd, st, dt, nl, nr) >> FAITH >>
OCP-COR(nt, ns, ld, rd, lt, rt, ls, rs)

Constraint Definitions

OCP-COR[αCONT][αVOICE] No adjacent coronals agreeing in voicing and continuancy
OCP-COR[αSON] No adjacent coronals agreeing in sonorancy
OCP-COR No adjacent coronals

The hierarchy (16) reflects the fact that while agreement in sonorancy is important in determining the strength of the co-occurrence restrictions in Muna, it is neither necessary nor sufficient to produce the strongest restriction, which obtains when segments agree in voicing and continuancy.

The upshot is that place co-occurrence patterns vary cross-linguistically in ways that are not predicted by the similarity metric, and that are consistent with the hypothesis that rankings of relativized OCP constraints can differ between languages. It should be noted, though, that the relativized OCP theory is likely not restrictive enough; ideally, a theory of the typology of place co-occurrence should allow more freedom than the similarity metric, but less than freely rankable relativized OCP constraints.

Frisch *et al.* (2004) derive the weakening of the co-occurrence restriction amongst coronals in Arabic from the large size of the Arabic coronal inventory relative to that of the labials and dorsals. The increased size of the inventory increases the number of natural classes that the segments being compared can belong to within their place of articulation, and hence lowers similarity values, since this number serves as the denominator in the similarity metric.

Table 5 presents the average similarity and O/E values for the pairs that can be compared across all three places of articulation (voiced and voiceless stops, nasals, and prenasalized stops), again omitting ones subject to the identity effect or the ban on multiple prenasals. The differences amongst the similarity values are relatively small; these would all be grouped together in the table comparing O/E and similarity in Frisch *et al.* (2004: 203). The O/E values, however, fall dramatically from the coronals to the labials.

	Coronals	Labials	Dorsals
O/E	0.78	0.22	0.09
Similarity	0.22	0.33	0.29

Table 5 O/E and average similarity values

The reason that the similarity metric fails to capture the strength of the coronal effect is that the coronal inventory in Muna is only marginally larger than that of the labials and the dorsal, as can be seen in the consonantal inventory in (15).[6] McCarthy (2003b) finds a similar problem in Rotuman: like Arabic, co-occurrence between coronals is much freer across sonorancy classes than within them, but unlike Arabic, the coronal inventory is not particularly large. The Rotuman and Muna facts suggest that the weakness of OCP effects between coronals is not simply a factor of inventory size. While no account of the coronal effect is offered here (cf. Alderete 1997), it cannot be considered an advantage of the similarity metric that it derives it from size of the coronal inventory.

[6] Muna also has a set of palatals recently borrowed from Indonesian, which are omitted from the inventory in (15). Van den Berg (1989: 16) states: “The palatal consonants /c/, /j/, and /y/ are marginal loan phonemes. The number of words containing these recent loan phonemes is very low.” He further notes that they are replaced in all but very recent loans.

In sum, the greater predictiveness of the similarity metric may be a liability, rather than an advantage, given that it does not allow for an account of the Muna data. A further concern with the similarity metric is that it is unclear how it could be extended to deal with the phonological alternations that arise in some languages as a response to OCP-PLACE violations (see section 4 for a Muna example). Since the present analysis of gradient phonotactics is cast within the OT model of generative phonology, which is constructed to produce alternations, this extension is of course straightforward.

3. Lexically specific constraints and learnability

The grammars that we have proposed for lexically gradient restrictions in Arabic and Muna raise three learnability problems (see also Ota 2004):

(17) i. How does a learner create lexically specific constraints for exceptions to phonotactics?
ii. How do the markedness constraints get in the right order?
iii. How do the faithfulness constraints get interspersed correctly?

Itô and Mester (1999) propose a solution to the second two of these problems, but it requires that the learner encounter the data in a particular order, and it does not address the first problem. Here we address this issue by proposing that learners are initially conservative, in that when they encounter a word that requires an adjustment to the grammar, they first assume that this adjustment is specific that word. More formally, in terms of Tesar and Smolensky (1998) et seq., when Error-Driven Constraint Demotion produces a Mark-Data pair, faithfulness constraints preferring the winner are indexed to the lexical item in question. When this proposal is incorporated into Prince and Tesar's (2004) Biased Constraint Demotion Algorithm, it automatically yields an answer to the second two problems.

3.1 The problem

In this section, we provide a simple hypothetical example of a language with a grammar incorporating lexically specific constraints, which will form the basis of the learnability discussion. The hypothetical language generally lacks both codas and clusters; most of the words are made up only of CV syllables. The lack of codas and clusters is a purely static generalization; there are no alternations. There is, however, a small set of words with codas (two words, [bat] and [net]), and an even smaller set with clusters (one word, [pla]).

(18) **Hypothetical grammar 1**
Grammar: MAX-L1 >> *COMPLEX >> MAX-L2 >> NOCODA >> MAX
Lexicon: /pla$_{L1}$/ /bat$_{L2}$/ /net$_{L2}$/ /pa/ /ti/ /la/ /me/ /go/ /ra/

MAX-L1 and MAX-L2 are lexically specific versions of MAX. The ranking of *COMPLEX and NOCODA above MAX enforces the general absence of clusters and codas. Since there are no alternations, the choice of this particular faithfulness constraint is arbitrary; clusters and codas could equally be avoided through epenthesis. The exceptional words /bat/ and /net/ are able to violate NOCODA because they indexed to MAX-L2, which ranks above the markedness constraint. Because it is indexed to MAX-L1, /pla/ is able to violate *COMPLEX.

The grammar in (18) encodes the three-way distinction shown in (19). In terms of the possible indexation approach to gradient acceptability, a CV nonce form would surface faithfully with any indexation (3/3), a CVC nonce form would surface faithfully if indexed to either MAX-L1 or MAX-L2, but not if left unindexed (2/3), and a CCV nonce form would surface faithfully only if indexed to MAX-L1, or in 1/3 indexations. In Coetzee's (2004) terms, CCV violates the highest ranked markedness constraint, CVC violates a lower ranked one, and CV violates none.

(19) **Ranking of word types by hypothetical stratified grammar**
CV > CVC > CCV

In Error-Driven Constraint Demotion (Tesar and Smolensky 1998), the learners' current grammar is used to parse the data it encounters. If the grammar yields an output that does not correspond to the learning data, the constraints are reranked using the Constraint Demotion Algorithm. Unmodified, this approach would yield the following grammar for our hypothetical language:

(20) **Hypothetical grammar 2**
Grammar: MAX >> *COMPLEX, NOCODA
Lexicon: /pla/ /bat/ /net/ /pa/ /ti/ /la/ /me/ /go/ /ra/

If there were a corresponding alternation that applied to regular but not exceptional forms, then inconsistency detection (Tesar 1998, Prince 2002) could be applied to trigger the creation of a lexically specific constraint.[7] But for purely phonotactic stratification, an unstratified grammar would successfully parse all of the forms. The problem is that the grammar in (20) is identical to one for a language that has no restriction on the occurrence of codas and clusters.

This is a special case of the subset problem for phonotactic learning discussed by Smolensky (1996), Hayes (2004) and Prince and Tesar (2004). These papers

[7] See Pater (2004) for some development of this approach; see also Ota (2004) for discussion of Japanese postnasal voicing in these terms.

posit learning biases in which markedness constraints are ranked over faithfulness constraints, so that learners are not trapped in the superset grammar with faithfulness over markedness. The twist here is that the language does in fact provide evidence for the F >> M ranking, but only in a limited set of words. To deal with this special case, we elaborate on Prince and Tesar's (2004) Biased Constraint Demotion (BCD) Algorithm (cf. Itô and Mester 1999, which elaborates on Smolensky 1996).

3.2 A solution

When Error-Driven Constraint Demotion detects an error, it creates a Mark-Data pair that provides the information used for constraint demotion. The learning datum is called the Winner, and the output of the learner's grammar is called the Loser. Constraints assign a 'W' when they prefer the Winner, and an 'L' when they prefer the Loser. We adopt Tesar's (1998) proposal that Mark-Data pairs are retained for further learning, forming a set that Tesar and Prince (2004) term a support. The main elaboration that we propose is that when Mark-Data pairs are formed, faithfulness constraints that prefer the winner are indexed to the lexical item in question.

BCD iteratively places constraints in strata according to the following steps, which favor high-ranking markedness constraints, and hence a restrictive grammar:

(21) i. Identify constraints that prefer no losers
ii. If any of these are markedness constraints, install them in the current stratum, and return to step i.
iii. If there are no available markedness constraints, install faithfulness constraints that prefer winners, and return to step i.
iv. If there are no faithfulness constraints that prefer winners, install those that prefer no losers, and return to step i.

To illustrate how this modified BCD algorithm creates a stratified grammar, we use the constraint set, and hypothetical language introduced in the last section:

(22) Words: [pa] [ti] [la] [me] [go] [ra] [mat] [net] [fle]
Constraints: NoCoda *Complex Max

Before any data are presented to the algorithm, it creates the following grammar, with the markedness constraints ranked above the faithfulness constraint:

(23) NoCoda, *Complex >> Max

If this grammar is used to parse any of the non-CV words, it will yield an error. For example, given [mat], the grammar yields [ma]. A Mark-Data pair is then created, with an indexed faithfulness constraint:

(24)

Input	W ~ L	NOCODA	*COMP	MAX-L1
mat_{L1}	mat ~ ma	L		W

BCD will now produce the following grammar:[8]

(25) *COMPLEX >> MAX-L1 >> NOCODA >> MAX

Since the indexed constraint applies only to /mat/, the grammar in (25) would also produce an error upon encountering [net] (and [fle]). All of the words with marked structures will lead to errors and the creation of Mark-Data pairs, while the unmarked CV words will never produce errors or Mark-Data pairs. The full support tableau for this language will thus be as in (26).

(26)

Input	W ~ L	NO CODA	*COMP	MAX-L1	MAX-L2	MAX-L3
mat_{L1}	mat ~ ma	L		W		
net_{L2}	net ~ ne	L			W	
fle_{L3}	fle ~ fe		L			W

Applying BCD to this support entails choosing amongst three lexically specific faithfulness constraints. Prince and Tesar (2004: 267) propose that the choice amongst faithfulness constraints is made by identifying ones that "free up" markedness constraints for ranking:

(27) "**Smallest effective F sets.** When placing faithfulness constraints into the hierarchy, place the *smallest set* of F constraints that *frees up some markedness constraint.*"

To free up a markedness constraint means to eliminate all its L marks. A Mark-Data pair is eliminated when a constraint that prefers its Winner is installed; the installation of that constraint guarantees that the Winner will be optimal in the resulting grammar. Installing MAX-L3 will eliminate the Mark-Data pair for /fle/, and will free up *COMPLEX. To free up NOCODA, both MAX-L1 and MAX-L2 must be installed. Therefore, the smallest effective F set is {MAX-L3}. When this

[8] There is one additional complication: the algorithm must prefer ranking the lexically specific MAX-L1 over the general MAX. This can be attributed to the preference for ranking specific over general versions of faithfulness constraints (see Smith 2000, Hayes 2004; cf. Prince and Tesar 2004).

constraint is installed, the Mark-Data pair for /fle/ is eliminated, indicated by removing the row from the support tableau in (28):

(28) Grammar: MAX-L3 >>

Input	W ~ L	NO CODA	*COMP	MAX-L1	MAX-L2	MAX-L3
mat_{L1}	mat ~ ma	L		W		
net_{L2}	net ~ ne	L			W	

*COMPLEX is then installed due to the markedness bias:

(29) MAX-L3 >> *COMPLEX >>

This does not eliminate any further Mark-Data pairs, since *COMPLEX prefers no Winners, so the support remains as in (28).

Before proceeding further, it is worth noting the important role of the "Smallest effective F sets" clause to the success of BCD in creating a stratified grammar. Part of the goal is to have markedness constraints ranked according to how often they are violated in the language, with higher rank correlating with fewer violations (see also Boersma 1998 in a different context). A lexically specific faithfulness constraint is created for each occurrence of a word with a violation of a markedness constraint. A markedness constraint that is violated rarely will create few faithfulness constraints. This will be a "small effective F set", and it, followed directly by this markedness constraint, will be installed before a markedness constraint that is violated more often since its effective F set will be larger. In the present example, *COMPLEX is only violated once, and so its associated effective F set consists only of MAX-L3. NOCODA is violated twice, so its effective F set consists of MAX-L1 and MAX-L2.

Once we have installed *COMPLEX, MAX-L1 and MAX-L2 will be installed together to free up NOCODA. This eliminates all of the Mark-Data pairs from further consideration, so the support tableau is now empty.

(30) MAX-L3 >> *COMPLEX >> MAX-L1, MAX-L2

With no more data to account for, NOCODA will be installed next due to the markedness bias, and then the general MAX constraint:

(31) MAX-L3 >> *COMPLEX >> MAX-L1, MAX-L2 >> NOCODA >> MAX

Lexically specific constraints can be collapsed as follows:

(32) Merge instantiations of any constraint that occupy the same stratum

This produces the desired grammar, and lexicon:

(33) Grammar: MAX-L1 >> *COMPLEX >> MAX-L2 >> NOCODA >> MAX
Lexicon: /fle_{L1}/ /mat_{L2}/ /net_{L2}/ /pa/ /ti/ /la/ /me/ /go/ /ra/

Further collapse of lexically specific constraints is necessary to produce a non-stratified grammar when a structure is well attested in a language. This can be accomplished by imposing a maximum size on the set of words targeted by a lexically specific constraint:

(34) If the number of words indexed by a constraint is greater than *x*, remove indexation, and delete any lower ranked instantiation of the constraint

Once indexation has been removed, the learner will also stop making errors, and creating Mark-Data pairs and lexically specific constraints. For example, if we assumed that $x = 1$,[9] the step in (34) would result in (35) for our hypothetical language.

(35) Grammar: MAX-L1 >> *COMPLEX >> MAX >> NOCODA
Lexicon: /fle_{L1}/ /mat/ /net/ /pa/ /ti/ /la/ /me/ /go/ /ra/

If a learner with this grammar encountered another word with a coda, it would parse it faithfully, and no Mark-Data pair would be created.

The arbitrary diacritics that lexical indexation creates may also be eliminated in favor of grammatical categories. For example, Smith's (1997) cross-linguistic study shows that nouns often contain structures that are banned in verbs. Learners of these languages might start by treating these structures as arbitrary exceptions, and then identify the category Noun as the property uniting the indexed items, which would replace the arbitrary diacritic (e.g. 'MAX-L1' → 'MAX-NOUN'). Since faithfulness constraints targeting categories do not require lexical markings, these may be more robust than arbitrary exceptions, and less prone to regularization (see also Anttila 2002 for evidence of an arbitrary pattern being grammaticalized). Similarly, strata like those discussed Itô and Mester (1999), which tend to show a clustering of properties, might use just one (arbitrary) lexical diacritic across a number of faithfulness constraints, which could also increase robustness.

[9] It is of course impossible to know exactly what value *x* should have, though 1 is clearly too low for real cases. The larger *x* is, the more well-formedness distinctions will be made the grammar.

3.3 Conclusions on learning

In this section, we have shown that a grammar encoding gradient phonotactics can be learned by making only one addition to the Biased Constraint Demotion Algorithm: that lexically specific faithfulness constraints are created when Mark-Data pairs are formed. With this one addition, the markedness bias and the smallest effective faithfulness sets clause place the constraints in the appropriate order. It is possible that Boersma's (1998) Gradual Learning Algorithm would also succeed at this task if modified in the same way, since it ranks constraints according the frequency with which they are violated. One issue that would seem a little less tractable using the Gradual Learning Algorithm is how constraints get collapsed (see (32) and (34)), since there is no notion of a stratum in that theory. Though we leave the full exploration of this possibility for further research, each algorithm would appear to have its own advantage related to this domain. The Constraint Demotion Algorithm detects inconsistency, which seems crucial in learning some cases of exceptionality (Tesar *et al.* 2003, Pater 2004, Tesar 2004), as well as for learning underlying forms (Tesar and Prince 2004, McCarthy 2004) and metrical structure (Tesar 1998; cf. Apoussidou and Boersma 2004). The Gradual Learning Algorithm does not detect inconsistency, but it does handle variation, and has been applied to gradient exceptionality in alternations (Zuraw 2000).

4. Lexically specific constraints and perception

We have argued for the inclusion of lexically specific constraints in OT on the basis of their ability to produce a grammar reflecting lexically gradient well-formedness. However, one might maintain that phonology is responsible only for alternations, or only for exceptionless lexical generalizations, and thus find this argument unpersuasive. The "alternation-only" stance would run contrary to the aim of Chomsky and Halle (1968) and subsequent generative phonologists to use a single system to capture phonological alternations and lexical patterns (that is, to solve the duplication problem). This goal is based on the observation that alternations often serve to bring words in conformity with the patterns present in the lexicon. Pater and Tessier (2003) provide experimental evidence that speakers do use their knowledge of static lexical generalizations in learning alternations, which supports the single system approach (see also Hayes 2004 and Tesar and Prince 2004 on the role of lexical patterns in the learnability of alternations).

A nearby example of place co-occurrence restrictions triggering alternations is found in the blocking of Muna /um/ infixation (36a,b) with labial-initial roots (36c-h) (van den Berg 1989, Pater 2001; see also Tessier 2004 for other cases):

(36)	a. /um+dadi/	[dumadi]	'live'
	b. /um+gaa/	[gumaa]	'marry'

c. /um+pili/	[mili]	'choose'
d. /um+futaa/	[mutaa]	'laugh'
e. /um+baru/	[baru]	'happy'
f. /um+bhala/	[bhala]	'big'
h. /um+waa/	[maa]	'give'

Phonological analysis has also not typically restricted its attention to exceptionless generalizations, presumably because of a belief that a small set of exceptions will not force a speaker to ignore an otherwise robust pattern. And it is certainly uncontroversial in generative phonology to make a distinction between a regular, an exceptional, and an impossible pattern, that is, to have a three-way gradation in acceptability. However, one might still balk at the necessity to encode further degrees of gradience in a phonological grammar.

The question of what distinctions should be encoded in a grammar is, we believe, ultimately an empirical one. It is again standard in generative phonology to appeal to native speaker judgments to justify the need for a grammar to rule out structures that are unattested in a language. As others have argued, if these judgments reveal that different structures are ranked along a graded scale, then the phonological grammar should reflect this gradience (see e.g. Zuraw 2000, Berent *et al.* 2001, Hayes and Boersma 2001, Frisch and Zawaydeh 2001, Albright and Hayes 2003, Coetzee 2004, and Hammond 2004 on gradient well-formedness judgments, as well as Pierrehumbert 2001 and references therein). In this section, we provide evidence that speakers do rank structures that violate different OCP-PLACE constraints in a way that corresponds to the degree to which those OCP-PLACE constraints are obeyed in the lexicon. We then show that the present approach to lexically gradient acceptability resolves a problem in attributing well-formedness judgments to the activity of the phonological grammar: that these judgments may be sensitive to cumulative constraint violation, in a way that phonological alternations do not seem to be.

4.1 Experimental backgound

English has a consonantal co-occurrence restriction that has some of the same key features as Arabic and Muna. In particular, its strength varies by place of articulation, with coronals co-occurring more freely than dorsals, and dorsals more freely than labials. The English restriction applies to consonants in words of the shape $sC_1V(C)C_2$: C_1 and C_2 are underrepresented if they are homorganic (see e.g. Fudge 1969, Davis 1991, Lamontagne 1993). However, the degree of underrepresentation varies by place: labial oral stops are unattested, dorsals are somewhat more common, and coronals are so common that they are sometimes taken to co-occur freely (cf. Lamontagne 1993: 267). Examples appear in (37).

(37) sTVT e.g. stud, stood, stead, stat, state, staid, stilt, stunt, stand…
sKVK e.g. skeg, skag, skunk, skank, skulk
sPVP e.g. ?

Here we will use cover labels for the constraints at issue (see Baertsch and Davis 2003 and Coetzee 2004 for analyses). 'T', 'P', and 'K' are taken to stand for oral stops at each place of articulation, and V is a placeholder for a vowel and optional sonorant consonant. The ranking of the constraints that is appropriate for English is shown in (38):

(38) *sPVP >> FAITH-L1 >> *sKVK >> FAITH-L2 >> *sTVT >> FAITH

This ranking rules out sPVP words, in accordance with their absence from the English lexicon, and gives sKVK a status intermediate between them and sTVT. It also encodes the assumption that sTVT words are less well-formed than sCVC words with heterorganic stops (see Coetzee in press).

To test whether English native speakers are sensitive to the three-way distinction between the word types sTVT, sKVK and sPVP, Coetzee (2004) examined well-formedness ratings and lexical decision times for nonce words.[10] This draws on the work of Berent *et al.* (2001), who used these measures to show that Hebrew speakers are sensitive to a three-way distinction between different types of $C_1VC_2VC_3$ root. Nonce roots with identical consonants in C_1 and C_2 were rated worse than ones with identical C_2 and C_3, and were recognized as non-words more quickly. In addition, nonce roots with no identical consonants were rated the best of all, though they were rejected as non-words more quickly than ones with identical consonants in C_2 and C_3 (see Coetzee 2004: 388 for discussion).

The stimuli for Coetzee's English experiment were nonce words of the shape sC_1VC_2, where C_1 and C_2 are identical voiceless stops, [p], [t], and [k] separated by just a vowel. They were selected to control for aspects of lexical frequency other than the ones at issue. Lexical neighborhoods and phoneme transitional probabilities have been shown to affect both well-formedness judgments and lexical decision time, with denser lexical neighborhoods and higher phoneme transitional probabilities resulting in higher well-formedness and slower lexical decision time (Newman, Sawusch and Luce 1997, Vitevich and Luce 1999). The stimuli for the word types that were hypothesized to be more grammatically well-formed had neighborhood densities and cumulative bi-phone probabilities that were equal to or lower than those for the word types that were hypothesized to be less

[10] Coetzee's (2004) hypothesis that there is a gradient distinction between the word types comes not from their relative lexical frequency, but from a universal ordering of *sPVP >> *sKVK >> *sTVT. If we consider words that are of the exact form of the stimuli in the experiment (only voiceless stops without intervening nasals or liquids), both sPVP and sKVK are unattested, so lexical frequency cannot directly explain the results. In this paper we take the position that the constraints at issue are broader in scope than the ones posited in Coetzee (2004), and that their ranking is affected by frequency.

well-formed, thus controlling for the effect of these factors. For example, the neighborhood densities and cumulative bi-phone probabilities of the sTVT words were equal to or lower than those of the sKVK words. In this context, it is also worth noting that there are in fact no real words of exactly the same shape as the sKVK stimuli: all existing sKVK words either end in a voiced stop (i.e. *skag*, *skeg*) or have a word final cluster (i.e. *skulk*, *skank*, *skunk*). Thus, if lexical statistics are responsible for the difference between them and sPVP, the influence must be somewhat indirect: our claim is that this influence is mediated by the system of phonological constraints.

Coetzee (2004) conducted three experiments with the same group of 20 subjects (undergraduates from the University of Massachusetts, Amherst): a gradient well-formedness task, a comparative well-formedness task, and a lexical decision task. All three experiments confirmed to the predictions of the hypothesized gradient well-formedness scale: sTVT > sKVK > sPVP, though in the gradient well-formedness task, the difference between sKVK and sPVP was not significant. Here we will discuss the details of only the second two tasks.

4.2 Comparative well-formedness

In this experiment, subjects were presented with non-word token pairs, and were asked to select from each pair the member that they considered to be most well-formed/most likely to be included in the lexicon of English. Each pair had one of the following forms: [sTVT]~[sKVK], [sTVT]~[sPVP], or [sKVK]~[sPVP]. The lefthand member of each pair is the one predicted to be judged as more well-formed according to the hierarchy in (38). There were 15 pairs of each form, and for every one, the word matching the shape of the righthand member had the higher cumulative biphone probability, and most instances, it also had a higher lexical neighborhood density. In this way, these lexical measures conflicted with the predictions of the grammar. The overall statistics for each condition are shown in (39).[11]

[11] The lexical statistics were all calculated from the CELEX database (Baayen *et al*., 1995). Lexical neighborhood density and cumulative bi-phone probabilities were calculated according to the method used by Vitevitch and Luce (1999) and Newman *et al*. (1997). The neighbors of a token are defined as any word that can be formed from the token by substitution, addition or deletion of one phoneme from the token. Lexical neighborhood density is then calculated as follows: (i) Find all the neighbors for a token. (ii) Sum the log frequencies of all the neighbors. To determine the cumulative bi-phone probablity for the bi-phone sequence XY, count the number of times that the XY sequence occur and take the log of this frequency. Count the total number of bi-phone sequences and take the log of this count. Now divide the log of XY by the log the total number of bi-phone sequences.

(39) **Lexical statistics for comparative well-formedness judgment experiment**
(LND = lexical neighborhood density; CBP = cumulative bi-phone probability)

a. T~K-condition

	[sTvT]	[sKvK]	
LND	16.55	24.21	$t(14) = 3.32$, one-tailed $p < .003$
CBP	0.17	0.26	$t(14) = 6.78$, one-tailed $p < .000$

b. T~P-condition

	[sTvT]	[sPvP]	
LND	14.87	21.45	$t(14) = 4.30$, one-tailed $p < .001$
CBP	0.16	0.25	$t(14) = 6.55$, one-tailed $p < .000$

c. K~P-condition

	[sKvK]	[sPvP]	
LND	10.30	21.26	$t(14) = 5.26$, one-tailed $p < .000$
CBP	0.15	0.25	$t(14) = 4.91$, one-tailed $p < .000$

Along with the 45 test-pairs, there were 45 non-word filler pairs. The words were produced by a phonetically trained female native speaker of American English, and were normalized for intensity. Subjects judged each of the pairs twice, with the order of the members of the pair reversed between the two presentations. Subjects were instructed that they would hear a pair of non-words, and that their task would be to select the member of each pair that they thought could most likely be included in the lexicon of English in the future. They indicated their response by pressing a button.

Figure 2 presents the mean number of choices in each condition by subject. In all cases, the word-type that is rated as more well-formed by the grammar in (38) is selected more frequently.

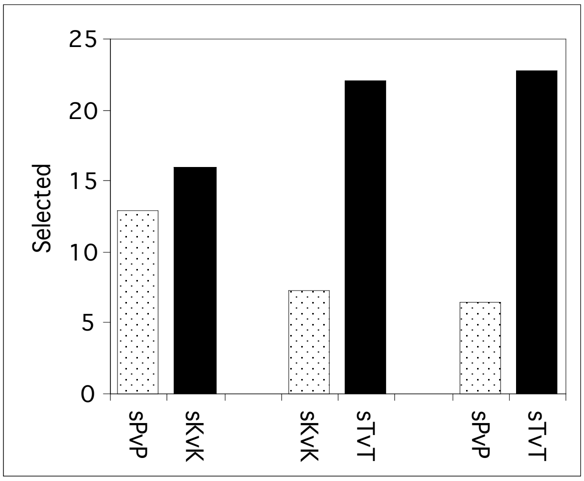

Figure 2: Results of the comparative well-formedness task

The results were subjected to a 2 × 3 ANOVA with hypothesized markedness (relatively marked~unmarked by the grammar in (38)) and condition (K~P, T~K, T~P) as independent variables. A main effect of markedness was found both by subjects ($F(1,19) = 23.28$, $p < 0.000$) and by items ($F(1,14) = 188.43$, $p < 0.000$). There was also a significant interaction between markedness and condition both by subjects ($F(2,18) = 10.37$, $p = 0.001$) and by items ($F(2,13) = 23.91$, $p < 0.000$).

The contrast between the marked and unmarked tokens in each condition was further investigated with one-tailed *t*-tests. In the K~P-condition there was an advantage for the less marked [sKvK]-tokens over the more marked [sPvP]-tokens. This difference was significant by items ($t(14) = 1.92$, $p = 0.037$), but not by subjects ($t(19) = 1.12$, $p = 0.14$). In the T~K-condition, the unmarked [sTvT]-tokens were preferred over the marked [sKvK]-tokens both by subjects ($t(19) = 4.54$, $p < 0.000$) and by items ($t(14) = 15.58$, $p < 0.000$). Similarly, in the T~P-condition, the unmarked [sTvT]-tokens were preferred over the marked [sPvP]-tokens, both by subjects ($t(19) = 5.73$, $p < 0.000$) and by items ($t(14) = 13.09$, $p < 0.000$).

The comparative well-formedness judgments thus correspond to the hypothesized gradient well-formedness scale: sTVT > sKVK > sPVP, although the evidence for the preference for sKVK over sPVP is somewhat less strong than the other distinctions.

4.3 Lexical decision

In this experiment, subjects are presented auditorily with a list of words and non-words, and are asked to decide as quickly as possible whether each word is a real word. The basic hypothesis is that listeners use, among other things, the information provided by grammar when they make lexical decisions. The less well-formed a non-word token is, the less seriously a listener will consider it as a possible word, and the quicker the token will be rejected. The non-words that the listeners were presented with included tokens of the form sTVT, sKvK and sPvP, and it was predicted that the speed of lexical decision should correlate negatively with their well-formedness: sPVP should be rejected the fastest, sTVT the slowest, and sKVK at an intermediate rate. In this case the tokens were selected such that the mean lexical neighborhood density and the mean cumulative bi-phone probability of the word forms being compared did not differ significantly. The mean lexical statistics of the tokens in each of the three conditions are given in (40).

(40) **Lexical statistics for lexical decision experiment**
(LND = lexical neighborhood density; CBP = cumulative bi-phone probability)

a. T~K-condition

	[sTvT]	[sKvK]	
LND	31.58	27.24	$t(8) = 0.78$, two-tailed $p > .45$.
CBP	0.28	0.24	$t(8) = 1.35$, two-tailed $p > .21$

b. T~P-condition

	[sTvT]	[sPvP]	
LND	31.58	26.56	$t(8) = 1.00$, two-tailed $p > .34$
CBP	0.28	0.22	$t(8) = 1.84$, two-tailed $p > .10$

c. K~P-condition

	[sKvK]	[sPvP]	
LND	14.95	14.26	$t(8) = 0.10$, two-tailed $p > .92$
CBP	0.18	0.16	$t(8) = 0.38$, two-tailed $p > .71$

In each condition there were 5 tokens for each word type, but because some of the tokens were used in more than one comparison there were a total of 27 test-tokens. Each of these 27 test-tokens was included once in the list. To this 76 fillers were added so that the final list contained 103 tokens. These tokens were presented auditorily to subjects. On each presentation the list was differently randomized. The list was presented twice to subjects, with a break of about five minutes between presentations. Subjects responded by pressing one of two buttons on a response box. One button was marked as “Yes”, and was used to indicate that the token was a word of English. The other button was marked as “No”, and was used to indicate that the token was not a word of English. The order between the buttons was varied so that half of the subjects responded “Yes” with the right hands, and half responded “No” with their right hands. Subjects were instructed to respond as quickly as possible, but to listen to the whole token before responding.

The response times were recorded starting at the onset of a stimulus. Before analysis, the duration of every stimulus was subtracted from the recorded response time. The resulting measure represented how long after (or before) the end of the stimulus a subject recorded a response. In the rest of the discussion I will refer to this measure (recorded response time minus token duration) as “response time”. The response times for each subject were normalized, and responses that were more than 2 standard deviations away from the mean for a subject were excluded from the analysis. Only correct non-word responses were included in the analysis. Exclusion of outliers and incorrect responses resulted in exclusion of only 8% of the total responses.

Figure 3 presents the mean reaction times in ms by subject for the lexical decision task.

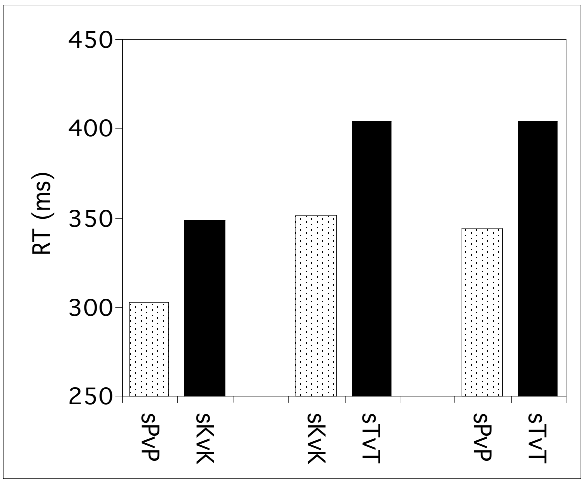

Figure 3: Results of the lexical decision task

The response time data were subjected to a 2 × 3 ANOVA with markedness (marked~unmarked) and condition (K~P, T~K, T~P) as independent variables. A main effect of markedness was found by subjects ($F(1, 19) = 21.68$, $p = 0.001$), but not by items ($F(1, 4) = 2.584$, $p = 0.18$). There was no interaction between markedness and condition by subjects ($F(2, 18) = 0.58$, $p = 0.57$) or by items ($F(2, 13) = 0.08$, $p = 0.92$).

The contrast between the marked and unmarked tokens in each condition was further investigated with one-tailed *t*-tests. In the K~P condition the more marked

[sPvP]-tokens had shorter reaction times than the less marked [sKvK]-tokens. This difference was significant both by subjects ($t(19) = 3.\ 79$, $p < 0.000$) and by items ($t(8) = 2.15$, $p = 0.03$). In the T~K-condition the more marked [sKvK]-tokens had shorter reaction times than the less marked [sTvT] tokens. This difference was significant by subjects ($t(19) = 3.40$, $p < 0.002$) but not by items ($t(8) = 1.63$, $p = 0.07$). In the T~P-condition the more marked [sPvP]-tokens were also rejected more quickly than the less marked [sTvT]-tokens. This difference was significant by subjects ($t(19) = 4.\ 20$, $p < 0.001$) but not by items ($t(8) = 1.55$, $p = 0.08$).

The results of the lexical decision task thus also confirm the predictions of the hierarchy in (38): the less well-formed a non-word is, the quicker it is rejected. While again not all of the comparisons were significant by items and by subjects, it is interesting to note that the word-type comparisons that are significant on both types of *t*-test differ in these two experiments: in this experiment it was [sPVP]~[sKVK], and in the last it was [sKVT]~[sTVT] and [sPVP]~[sTVT]. This suggests that the lack of significance is a result of experimental factors, rather than due to one of the distinctions being less robust.

In sum, the results of these two experiments provide strong evidence of native speaker knowledge of gradient restrictions based on lexical frequency, and thus motivate their inclusion in phonological grammars (see also Gelbart 2005 on the role that listeners' knowledge of correlations between exceptional restrictions plays in perception).

4.4 Cumulative constraint violation in acceptability

Hay, Pierrehumbert and Beckman (2001) present results that they claim to show that well-formedness judgments are sensitive to cumulative constraint violation (see further Pierrehumbert 2001). In the relevant experiment, subjects were asked to judge a variety of nasal-obstruent sequences for whether they spanned a word boundary; clusters that were more marked were hypothesized to be more frequently judged as heteromorphemic. One finding that emerged was that clusters containing a strident fricative were more often interpreted as heteromorphemic when a strident was present in the onset of the preceding syllable. They interpret this as the cumulative effect of an OCP constraint on stridents and the constraints on nasal-obstruent sequences. Because their experiment was not designed to test for this effect, they did not have a baseline condition for the OCP violation alone. However, the fact that there were differences between clusters with the same strident, but different nasals (e.g. [mʃ] vs. [nʃ]) suggests that it is unlikely that the OCP itself explains the judgments. Coleman and Pierrehumbert (1997) point out that such cumulative constraint activity is unexpected in standard generative phonology, or standard Optimality Theory.[12]

[12] An alternative explanation for these experimental results is that the presence of multiple ill-formed structures increases the likelihood that subjects will notice or respond to any one of them. It seems

However, the approach to gradient acceptability proposed here does in fact predict cumulative constraint activity, as we will now show by using the possible indexation approach to analyze another simple hypothetical example. In this language, onsetless syllable and codas are generally banned, but there are a few exceptional words containing instances of each. The grammar for this language would contain rankings like those shown in (41):

(41) C-MAX-L1 >> NOCODA >> C-MAX
V-MAX-L2 >> ONSET >> V-MAX

The repair for NOCODA violations is assumed to be consonant deletion (e.g. /bat/ -> [ba]), and for ONSET vowel deletion (e.g. /uma/ -> [ma]). These violate Max constraints relativized to consonants (C-MAX) and vowels respectively (V-MAX). The exceptional words with codas and onsetless syllables are targeted by lexically specific versions of these constraints.

A word could be specified as L1, as L2, as both L1 and L2, or as neither. Given this set of possible indexations, the outcomes for a nonce word containing just a coda, just an onsetless syllable, or both a coda and an onset syllable are shown in (42).

(42)

	a.	*bat	✓bat_{L1}	*bat_{L2}	✓$\text{bat}_{L1,L2}$	2/4
	b.	*aba	*aba_{L1}	✓aba_{L2}	✓$\text{bat}_{L1,L2}$	2/4
	c.	*abat	*abat_{L1}	*abat_{L2}	✓$\text{abat}_{L1,L2}$	1/4
	d.	bat, aba > abat				

A word with both a coda and an onsetless syllable (42c) will surface faithfully only if specified as both L1 and L2. It is thus rated as less well-formed than a word with just one of the marked structures (42a, b), which will surface faithfully under two of the possible indexations.

An important attribute of this account of the cumulative effect of constraint violation in well-formedness judgments is that it is compatible with the absence of such cumulative effects in phonological alternations. The cumulative effects occur as a by-product of how gradient acceptability is calculated, but cannot occur in alternations because of OT's tenet of strict constraint domination. No ranking of ONSET, NOCODA, V-MAX and C-MAX will yield deletion of codas only from onsetless syllables, or vowels only from onsetless syllables with codas. This is in line with apparent absence of such alternations cross-linguistically, despite the large number of languages that delete consonants from coda position, and vowels in

---

difficult to tease these explanations apart, but doing so would be crucial for the evidence of cumulative constraint activity in well-formedness judgments to be fully convincing.

hiatus.[13] Unrestricted cumulative constraint activity can create even more bizarre results, such as vowel reduction in the presence of a complex onset anywhere in the word, or cluster reduction in all syllables when one has a voiced coda. One of the most striking results of the typological research that has been conducted in Optimality Theory is that there seems to be very little counter-evidence for strict domination. Local conjunction (Smolensky 1995) does create the effect of cumulative constraint activity, but the main objection to this elaboration of OT is that it is overly powerful (see e.g. Padgett 2002, McCarthy 2003a, Kawahara 2005).

It is an important question for further research to determine the extent to which the cumulative constraint activity observed in grammaticality judgments matches that found in both in lexical restrictions and in phonological alternations. The approach to gradient well-formedness judgments proposed here provides one way of explaining their divergences from the more traditional data sources for phonological analysis. Another issue for further research in this domain is the manner in which lexical frequency is linked to acceptability judgments. For example, do speakers base their judgments on O/E values, as assumed in section 1, and in the work of Pierrehumbert and colleagues, or on frequency of constraint violation, as assumed in the learnability account in section 3 (though see fn 3), and in Boersma and Hayes' (2001) analysis of variation and related well-formedness judgments?

5. Conclusions

In this paper we have shown that a version of OT with lexically specific constraints can produce grammars that reflect gradient patterns in the lexicon, and that these grammars can be learned with a minimal elaboration of Prince and Tesar's (2004) Biased Constraint Demotion Algorithm. This allows for an analysis of some traditionally problematic data on Arabic place co-occurrence. Frisch *et al.*'s (2004) critique of an OT analysis of place co-occurrence loses force in light of the possibility of accounting for gradient phonotactics in OT, and also in view of the difficulties that their alternative analysis has in accounting for the place co-occurrence patterns in Muna. Support for the incorporation of lexically based gradience into the phonological grammar comes from experimental evidence that speakers do rank structures according to the degree to which they are lexically attested.

---

[13] Levelt and van de Vijver (2004) claim that a restriction against specifically VC syllables, and not CVC or V, holds in a stage of child language. However, this stage is derived from a statistical analysis of a group of children's productions, and they do not provide evidence of a stable stage in one child's productions in which only target VC syllables are altered. They also mention Central Sentani (Hartzler 1976) as a possible example of a language that has this same restriction, but again, no examples of the application of a process affecting only VC syllables are provided.

References

Albright, Adam and Bruce Hayes. 2003. Rules vs. analogy in English past tenses: a computational/experimental study. *Cognition* 90: 119-161.

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.

Anttila, Arto. 1997. Deriving variation from grammar. In F. Hinskens, R van Hout and W. L. Wetzels (eds.) Variation, Change and Phonological Theory. Amsterdam, John Benjamins.

Anttila, Arto. 2002. Morphologically Conditioned Phonological Alternations. Natural Language and Linguistic Theory 20. 1-42.

Apoussidou, Diana and Paul Boersma. 2004. Comparing two Optimality-Theoretic learning algorithms for Latin stress. In B. Schmeiser, V. Chand, A. Kelleher and A. Rodriguez (eds.) *WCCFL 23 Proceedings*. Somerville, MA: Cascadilla Press. 101-114.

Baayen, R., Piepenbrock, R., and Gulikers, L. 1995. *The CELEX lexical database (CD-ROM)*. Philadelphia, PA: Linguistic Data Consortium.

Baertsch, Karen and Stuart Davis. 2003. The Split Margin Approach to Syllable Structure. In T.A. Hall and Silke Hamaan (eds.) *ZAS Papers in Linguistics* 32: 1-14.

Berent, Iris, Everett, Daniel L., and Shimron, Joseph. 2001. Do phonological representations specify variables? Evidence from the Obligatory Contour Principle. *Cognitive Psychology* 42. 1-60.

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

Berg, René van den and La Ode Sidu. 1996 Muna-English dictionary. KITLV Press, the Netherlands.

Berkley, Deborah. 2000. *Gradient Obligatory Contour Principle Effects*. Ph.D. Dissertation, Northwestern University.

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.

Caballero Hernández, Gabriela. 2005. The Stress System Of Central Raramuri: Root Privilege, Prosodic Faithfulness And Markedness Reversals. Ms, University of California, Berkeley.

Chomsky, Noam, and Morris Halle. 1968. *The sound pattern of English*. New York: Harper and Row.

Coetzee, Andries. 2004. *What It Means To Be A Loser: Non-Optimal Candidates In Optimality Theory*. Ph.D. Dissertation, University of Massachusetts, Amherst.

Coetzee, Andries. In press. The Obligatory Contour Principle in the perception of English. In Sonia Frota, Marina Vigario and Maria João Freitas, eds.

Prosodies. Selected Papers from the Phonetics and Phonology in Iberia Conference, 2003. New York: Mouton de Gruyter.

Coetzee, Andries and Joe Pater. 2005. Gradient Phonotactics in Muna and Optimality Theory. Ms. University of Michigan and University of Massachusetts, Amherst.

Coleman John and Janet Pierrehumbert. 1997. Stochastic phonological grammars and acceptability. SIGPHON 1997.

Cowan, J. 1979. *Hans Wehr: A Dictionary of Modern Written Arabic*. Wiesbaden, Germany: Otto Harrasowitz.

Davis, Stuart. 1991. Coronals and the phonotactics of non-adjacent coronals in English. In C. Paradis, J.-F. Prunet (eds.) The Special Status of Coronals: Internal and External Evidence. San Diego: Academic Press. 49-60.

Dresher. B. Elan. 1989. Comments on "Guttural Phonology", paper presented at MIT Conference on Feature Geometry.

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 Zawadyeh. 2001. The Psychological Reality of OCP-Place in Arabic. *Language* 77. 91-106.

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

Fudge, E.C. 1969. Syllables. *Journal of Linguistics* 5:253-286.

Fukuzawa, Haruka. 1999. Theoretical Implications of OCP Effects on Features in Optimality Theory. Doctoral dissertation, University of Maryland, College Park.

Gafos, Adamantios. 1998. Eliminating Long Distance Consonantal Spreading, *Natural Language and Linguistic Theory* 16:2. 223-278.

Gelbart, Ben. 2005. *The Role of Foreignness in Phonology and Speech Perception.* Ph.D. dissertation, University of Massachusetts, Amherst.

Greenberg, Joseph H. 1950. The patterning of root morphemes in Semitic. *Word* 6.162-181.

Hammond, Michael. 1995. There is no lexicon! Ms, University of Arizona.

Hammond, Michael. 2004. Gradience, Phonotactics, and the Lexicon in English Phonology. *International Journal of English Studies.* 1-24.

Hartzler, M. 1976. Central Sentani Phonology. *Irian* 5. 66-81.

Hayes, Bruce. 2004. Phonological acquisition in Optimality Theory: the Early Stages. In R. Kager, J. Pater and W. Zonneveld (eds.), *Constraints in Phonological Acquisition. Cambridge University Press*. 158-203.

Inkelas, Sharon, Cemil Orhan Orgun and Cheryl Zoll. 1997. Implications of lexical exceptions for the nature of grammar. In I. Roca (ed.) *Constraints and derivations in phonology*. Oxford: Clarendon Press. 393-418.

Inkelas, Sharon and Cheryl Zoll. 2003. Is Grammar Dependence Real? Ms, UC Berkeley and MIT.

Itô, Junko and Armin Mester. 1995. Japanese phonology. In John Goldsmith, ed. The Handbook of Phonological Theory. Blackwell, Oxford. 817-38.

Itô, Junko and Armin Mester. 1999. The Phonological Lexicon. In N. Tsujimura (ed.) *Handbook of Japanese Linguistics*. Oxford: Blackwell. 62-100.

Itô, Junko and Armin Mester. 2001. Covert generalizations in Optimality Theory: the role of stratal faithfulness constraints. *Studies in Phonetics, Phonology, and Morphology* 7. 273-299.

Kawahara, Shigeto. 2005. A faithfulness scale projected from phonetic perceptibility: The case of voicing in Japanese. Ms., University of Massachusetts, Amherst.

Kraska-Szlenk, Iwona.1997. Exceptions in Phonological Theory. Proceedings of the 16th International Congress of Linguists. Pergamon, Oxford. Paper No. 0173.

Lamontagne, Greg. 1993. *Syllabification and Consonant Cooccurrence Conditions*. Doctoral Dissertation, University of Massachusetts, Amherst.

Levelt, Clara and Ruben van de Vijver. 2004. Syllable types in cross-linguistic and developmental grammars. In R. Kager, J. Pater and W. Zonneveld (eds.), *Constraints in Phonological Acquisition*. Cambridge University Press. 204-218.

McCarthy, John. 1981. A Prosodic Theory of Nonconcatenative Morphology, *Linguistic Inquiry* 12. 373–418.

McCarthy, John. 1988. Feature geometry and dependency: A review. *Phonetica* 45. 84–108.

McCarthy, John. 1994. The phonetics and phonology of Semitic pharyngeals. In *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology 3*, ed. Patricia Keating, 191-233. Cambridge: Cambridge University Press.

McCarthy, John. 2003a. Comparative markedness. *Theoretical Linguistics* 29, 1-51

McCarthy, John. 2003b. Rotuman consonant cooccurrence restrictions. Course handout, Ling 730, University of Massachusetts, Amherst.

McCarthy, John. 2004. Taking a Free Ride in Morphophonemic Learning. Ms, University of Massachusetts, Amherst. (To appear in the *Catalan Journal of Linguistics*).

McCarthy, John and Alan Prince. 1993. Generalized Alignment. *Yearbook of Morphology.* 79–153.

McCarthy, John and Alan Prince. 1999. Faithfulness and identity in prosodic morphology. In R. Kager, H. van der Hulst and W. Zonneveld (eds.) The Prosody-Morphology Interface. Cambridge: Cambridge University Press. 218-309.

Mester, R. Armin. 1986. Studies in Tier Structure. Doctoral dissertation, University of Massachusetts, Amherst.

Newman, Rochelle S., Sawusch, James R., and Luce, Paul A. 1997. Lexical neighborhood effects in phonetic processing. *Journal of Experimental Psychology* 23:873-889.

Nouveau, Dominique. 1994. *Language Acquisition, Metrical Theory, and Optimality: A Study of Dutch Word Stress*. PhD Dissertation, Utrecht University.

Ota, Mits. 2004 The learnability of the stratified phonological lexicon. *Journal of Japanese Linguistics* 20, 4.

Padgett, Jaye. 1995. *Stricture in Feature Geometry*. Stanford: Center for the Study of Language and Information - Dissertations in Linguistics.

Padgett, Jaye. 2002. Constraint conjunction versus grounded constraint subhierarchies in Optimality Theory. Ms, University of California, Santa Cruz.

Pater, Joe. 2000. Nonuniformity in English stress: the role of ranked and lexically specific constraints. Phonology 17. 237-274.

Pater, Joe. 2001. Austronesian nasal substitution revisited. In L. Lombardi, (ed.) Segmental phonology in Optimality Theory: Constraints and Representations. Cambridge University Press. 159-182.

Pater, Joe. 2004. Exceptions in Optimality Theory: Typology and Learnability. Presented at the Conference on Redefining Elicitation: Novel Data in Phonological Theory, New York University. (http://people.umass.edu/pater/exceptions.pdf)

Pater, Joe. 2005. Exceptionality as constraint indexation. Ms, University of Massachusetts, Amherst.

Pater, Joe and Anne-Michelle Tessier. 2003. Phonotactic Knowledge and the Acquisition of Alternations. In M.J. Solé, D. Recasens, and J. Romero (eds.) *Proceedings of the 15th International Congress on Phonetic Sciences*, Barcelona. 1177-1180.

Pierrehumbert, Janet. 1993. Dissimilarity in the Arabic verbal roots. In A. Schafer, ed. *NELS 23: Proceedings of the North East Linguistic Society*. Amherst: GLSA. 367-381.

Pierrehumbert, Janet. 2001. Stochastic Phonology. *Glot International* 5/6. 195-207.

Prince, Alan. 2002. Entailed Ranking Arguments. Ms., Rutgers University.

Prince, Alan, and Paul Smolensky. 1993/2004. *Optimality Theory: Constraint interaction in generative grammar*. Technical Report, Rutgers University and University of Colorado at Boulder, 1993. Revised version published by Blackwell, 2004.

Prince, Alan, and Bruce Tesar. 2004. Learning Phonotactic Distributions. In R. Kager, J. Pater and W. Zonneveld (eds.) *Constraints in Phonological Acquisition*. CUP. 245-291.

Rose, Sharon. 2000. Rethinking geminates, long-distance geminates and the OCP. *Linguistic Inquiry* 31. 85-122.

Selkirk, Elisabeth. 1991. Vowel Height Features: Evidence for Privativity and Dependency. Paper presented at UQAM, Montréal.

Smith, Jennifer. 1997. Noun faithfulness: On the privileged behavior of nouns in phonology. Ms., University of Massachusetts.

Smith, Jennifer. 2000. Positional faithfulness and learnability in Optimality Theory. In Rebecca Daly and Anastasia Riehl, eds. *Proceedings of ESCOL 99*. Ithaca: CLC Publications. 203-214.

Smolensky, Paul. 1995. On the Internal Structure of the Constraint Component *Con* of UG. Talk presented at UCLA, April 7, 1995.

Smolensky, Paul. 1996. The Initial State and Richness of the Base in Optimality Theory. Technical Report JHU-CogSci-96-4, Cognitive Science Department, Johns Hopkins University.

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 P. Tamanji and K. Kusumoto (eds.) Proceedings of the North East Linguistic Society 28. Amherst, MA: GLSA, University of Massachusetts. 469-483.

Tesar, Bruce, and Alan Prince. 2004. Using Phonotactics to Learn Phonological Alternations. In The Proceedings of CLS 39, Vol. II: The Panels.

Tesar, Bruce, Alderete, John, Horwood, Graham, Merchant, Nazarre, Nishitani, Koichi, and Prince, Alan. 2003. Surgery in language learning. In *Proceedings of the Twenty-Second West Coast Conference on Formal Linguistics*, ed. by G. Garding and M. Tsujimura, 477-490. Somerville, MA: Cascadilla Press.

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

Tessier, Anne-Michelle. 2004. Root-restricted markedness and morpho-phonological domains. Paper presented at *Montreal-Ottawa-Toronto Phonology Workshop, February 2004, University of Ottawa*.

Uhlenbeck, E.M. 1949. *De structuur van het Javaanse morpheem*. Bandung: Nix.

Ussishkin, Adam. 1999. The inadequacy of the consonantal root: Modern Hebrew denominal verbs and Output-Output correspondence. *Phonology* 16.3. 401-442.

Vitevitch, Michael S., and Luce, Paul A. 1999. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language* 40:374-408.

Yip, Moira. 1989. Feature Geometry and Co-occurrence Restrictions. *Phonology* 6. 349-374.

Zamma, Hideki. Predicting Varieties: Partial Orderings in English Stress Assignment. Ms, Kobe City University of Foreign Studies/University College London

Zuraw, Kie. 2000. *Patterned Exceptions in Phonology*. Ph.D. Dissertation, UCLA.