This brief paper resumes the discussion of gradient phonotactics initiated by Frisch, Pierrehumbert and Broe (2004), and continued in two more recent publications in this journal (Coetzee and Pater 2008, Anttila 2008). The aim is to highlight some of the challenges that gradient phonotactics raise for phonological theory, and to show how they can be addressed with numerically weighted phonological constraints (following Coetzee and Pater 2008 and Hayes and Wilson 2008).

I start with some definitions. In recent usage, phonotactics refer very generally to all restrictions on the structure of phonological representations. Under this meaning, they contrast with alternations, which refer to changes in the phonological shape of morphemes. Gradience in this context refers to greater than binary differences in well-formedness amongst representations (rather than to gradience in the sense of representations using continua). Gradient phonotactics cannot be expressed by theories of grammar that admit only a categorical distinction between ill-formed and well-formed representations for a language.

In its original formulation, Optimality Theory (OT; Prince and Smolensky 1993/2004: p. 4) is in this way a categorical theory of grammar: even though its constraints can be violated in well-formed representations, it makes no distinctions amongst grades of well-formedness. Coetzee and Pater (2008) discuss proposed extensions of OT to gradient phonotactics, including those in their own earlier work (Coetzee 2004, 2008ab, Pater 2005, and Coetzee and Pater 2006). They conclude that “there seems not yet to be a satisfactory account of gradient phonotactic acceptability available within [OT]”.

Anttila (2008) cites Coetzee and Pater’s conclusion, and challenges it by introducing a novel approach to gradient phonotactics in OT that he calls the Complexity Hypothesis. Here, I highlight what seem to be significant advances in the treatment of gradient phonotactics permitted by the approach advocated by Coetzee and Pater (2008), which is cast within the weighted constraints framework of Harmonic Grammar (HG: Legendre, Miyata, and Smolensky 1990, Smolensky and Legendre 2006). Insofar as similar advances have not been obtained within OT, under the Complexity Hypothesis or any other approach to gradient phonotactics, this draws into question Antilla’s own conclusion that “gradient phonotactics does not require numerically weighted constraints”.

* Thanks to Arto Antilla, Andries Coetzee and Colin Wilson for comments on an earlier draft of this paper, and to Karen Jesney for helpful discussion.

1 The traditional definition, which includes only restrictions on sequences of phones, is used in Anttila (2008). However, the issues under consideration relate to the broader meaning, which is the one assumed by Coetzee and Pater (2008), as well as much other recent work on phonotactics (e.g. Hayes 2004, Prince and Tesar 2004, Hayes and Wilson 2008).
1. Some challenges of gradient phonotactics

If we take grammatical theory to be supplying models of knowledge of language, then the fundamental challenge raised by gradient phonotactics seems to be the following:

(1) To model speakers’ knowledge of degrees of acceptability of phonological structures, including distinctions that are greater than binary.

As well as providing an account of speakers’ knowledge of gradient phonotactics, a second challenge is showing how this knowledge could be acquired. Before moving on to a discussion of how these challenges can be met in HG, as well as a brief comparison with the Complexity Hypothesis, I will first discuss how an OT grammar measures categorical well-formedness, and introduce the phonotactic learning problem. Let us imagine two languages, one that permits voiced obstruent codas (‘Voico-y’ for “voiced codas - yes”), and another that bans them (‘Voico-n’ for “voiced codas - no”). In OT, these would be characterized by two different constraint rankings: ‘Voico-n’ has a constraint against voiced codas ranked above a faithfulness constraint, and ‘Voico-y’ has the reverse ranking. For concreteness, I adopt the constraints in (2). VOICE→ONSET is a positional licensing constraint (e.g. Ito 1986, Goldsmith 1990, Lombardi 1991; cf. Lombardi 1999 and section 2 below for an alternative using positional faithfulness), and IDENT-VOICE is due to McCarthy and Prince (1999). I assume that the feature borne by voiced obstruents is [voice]. For these and all further constraints in the paper, a violation mark is incurred for each minimal structure that fails to meet its requirements (e.g. VOICE→ONSET assigns one violation mark for each feature [voice] that is not associated with a segment in onset position, IDENT-VOICE assigns a violation mark for each correspondent pair in which one and only one member is specified as [voice]).

(2) VOICE→ONSET
Every feature [voice] is associated with a segment in onset position

IDENT-VOICE
Every pair of segments in correspondence agrees in specification for [voice]

In OT, as in other theories of generative phonology, a surface representation (e.g. [tad]) is well-formed iff some underlying representation maps to it (e.g. /tad/ → [tad]). In OT theories of phonotactic learning (e.g. Smolensky 1996, Tesar and Smolensky 2000, Hayes 2004, Prince and Tesar 2004), a learner presented with a (non-alternating) surface representation takes the identity map as the underlying representation, and determines whether the grammar maps it back to the original surface representation (if it fails to do so, learning is then triggered). This is presumably more generally the means by which an OT grammar generates well-formedness ratings for surface forms (see Coetzee 2004 for development of this approach). In the candidate cells of the tableaux in (3), I use capital ‘Y’ and ‘N’ to indicate the targets and results of acceptability ratings: ‘N’ mark a representation classified as ill-formed, and ‘Y’ indicates a representation classified as well-formed.
A speaker of Voico-y would thus judge [tad] as ill-formed as in the tableau on the left. A Voico-y speaker, on the other hand, would judge [tad] as well-formed, as shown in the tableau on the right. The challenge of phonotactic learning arises for a learner of Voico-n: to find a grammar that maps /tad/ to something other than [tad], without positive evidence of such maps. The just-cited theories help the learner meet this challenge by equipping it with biases that result in markedness constraints, like \( \text{VOICE} \rightarrow \text{ONSET} \), being ranked above faithfulness constraints, like \( \text{IDENT-VOICE} \), in the absence of evidence to the contrary.

We can now move to one of the specific challenges posed by gradient phonotactics. Phonological analyses often distinguish between ‘exceptional’ structures and ones that are ‘regular’, but the underlying theories do not usually provide guidelines for making these distinctions, either for an analyst or a learner. As a hypothetical example of phonotactic exceptionality, let us take the case of a third language, ‘Voico-e’, which has just one word with a voiced obstruent coda. In the original version of OT, this language would have exactly the same ranking as Voico-y (Prince and Smolensky 1993/2004 do not offer a theory of exceptions to phonotactics). Thus, OT does not generate a difference between well-formedness ratings of [tad] by speakers of Voico-y vs. Voico-e. Since these are hypothetical languages, I can of course offer no concrete evidence that speakers of these languages do treat a word like [tad] differently, but the experimental literature on gradient phonotactics reviewed in Coetzee and Pater (2008) strongly suggests that they would.

A related challenge of gradient phonotactics is that they require a means of distinguishing between the well-formedness of structures within a given language. Remaining still in the simplified realm of the hypothetical, we can further imagine that Voico-e has a well-attested contrast between plain and palatalized stops in onset position. I assume that the feature marking the palatalized stops is [palatal], and that the following three constraints are at issue:

\[
\begin{align*}
(4) \quad & \text{PALATAL} \rightarrow \text{ONSET} \\
& \text{Every feature [palatal] is associated with a segment in onset position} \\
& \text{*PALATAL} \\
& \text{No [palatal] features} \\
& \text{IDENT-PALATAL} \\
& \text{Every pair of segments in correspondence agrees in specification for [voice]} \\
\end{align*}
\]

The OT ranking for these constraints for our hypothetical language is shown in (5). This ranking accepts as well-formed a palatal in onset position, which violates only *PALATAL, but rejects as ill-formed a representation with a palatal in coda position, which contravenes the requirement that the palatal feature be associated with a segment in onset position. Again, we see an
illustration of the challenge of phonotactic learning: the learner of Voico-e must place the faithfulness constraint between the two markedness constraints, even though the observed data would allow it to be at the top of the hierarchy.

(5) Rating of [p'at] and [tap] by the OT grammar for Voico-e

<table>
<thead>
<tr>
<th></th>
<th>PALATAL →ONSET</th>
<th>IDENT-PALATAL</th>
<th>*PALATAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y☞[p'at]</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>[pat]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>PALATAL →ONSET</th>
<th>IDENT-PALATAL</th>
<th>*PALATAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>N [tap]</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>☞[tap]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The following tableau shows the ranking of the full set of constraints relevant for voicing, including the context-free *VOICE constraint. Since voicing is permitted outside of onset position, IDENT-VOICE must dominate both markedness constraints.

(6) Rating of [tad] by the OT grammar for Voico-e

<table>
<thead>
<tr>
<th></th>
<th>IDENT-VOICE</th>
<th>VOICE →ONSET</th>
<th>*VOICE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y☞[tad]</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>[tat]</td>
<td></td>
<td></td>
<td>!</td>
</tr>
</tbody>
</table>

Even though there is only one word with a voiced coda in this language, the rating of a form with a voiced coda like [tad] is identical to that for a form with a palatal stop, [p'at], which is a structure that we are imagining is well-attested.

Since the original version of OT provides no means by which the relative attestedness of two unrelated structures in a language could affect their relative acceptability, we can conclude that it would also fail to model this category of gradient phonotactics. In the next section, I will show how Coetzee and Pater’s (2008) proposal allows for an account of speakers’ knowledge of these types of gradient well-formedness, as well as an account of how the constraint weights required for this analysis might be acquired.

2. Meeting challenges of gradient phonotactics in Harmonic Grammar

In HG, the harmony of a representation is calculated as the sum of its weighted constraint scores: each constraint violation is multiplied by that constraint’s weight, and the results are summed (see Smolensky and Legendre 2006 and Pater 2008ab for introductions to HG and references to research in that framework). The optimal candidate in an OT-like version of HG is the one with
the highest numerical harmony. Coetzee and Pater (2008) propose that acceptability is a candidate’s harmony minus that of the best candidate in the rest of its candidate set.²

This is illustrated in the tableaux for [tad] in (7). Weights are shown in the cells with the constraint names, and violation counts are indicated as negative integers. A candidate’s harmony appears at the end of its row; the harmony for /tad/ → [tad] in the left-hand tableau is 2⋅(−1) = −2. Acceptability is shown in the cell with the target of the rating; the acceptability of the same form is (−2) − (−1) = −1. As this example shows, sub-optimal or ill-formed representations receive negative acceptability scores, since their Harmony is less than the optima in their tableaux. These negative scores replace the Ns in the OT tableaux above. Optimal forms receive positive scores, since their Harmony is greater than that of their competitors; these positive scores replace Ys.

(7) Ratings of [tad] by two HG grammars

<table>
<thead>
<tr>
<th></th>
<th>VOICE → ONSET</th>
<th>IDENT-VOICE</th>
</tr>
</thead>
<tbody>
<tr>
<td>/tad/</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>[tad]</td>
<td>−1</td>
<td>−2</td>
</tr>
<tr>
<td>[tat]</td>
<td>1</td>
<td>−1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>VOICE → ONSET</th>
<th>IDENT-VOICE</th>
</tr>
</thead>
<tbody>
<tr>
<td>/tad/</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>[tad]</td>
<td>−1</td>
<td>−1</td>
</tr>
</tbody>
</table>

Like the original version of OT, the HG account thus distinguishes between the rating of [tad] by a speaker of Voico-n (−1) vs. the rating of [tad] by a speaker of Voico-y (+1). Although the weights in (7) are partially arbitrary, any correct grammar for a language with voiced coda obstruents will have a weight of IDENT-VOICE that is greater than the weight for VOICE→ONSET, and [tad] will therefore have a positive acceptability score, as in the tableau on the right. The outcome for the tableau on the left requires a greater weight for VOICE→ONSET than for IDENT-VOICE. As there is no positive evidence for this strict inequality, this is the HG version of the OT phonotactic learning problem.

The ability of HG to model syntactic gradient well-formedness was its main original motivation (Legendre, Miyata and Smolensky 1990). Because the acceptability scores in the present HG approach to phonotactics are on a continuum, it can similarly encode gradient well-formedness (see also Hayes and Wilson 2008 for an earlier use of a harmony-based measure of phonotactic well-formedness). For the simple hypothetical example introduced in the last section, we can say that the acceptability of a voiced coda in the language Voico-y is greater than the acceptability of a voiced coda in Voico-e, the language in which only one word has such a structure. The tableau in (8) shows a constraint weighting that allows voiced codas to surface, but by a narrower margin than if IDENT-VOICE had a higher value, as in the right-hand tableau in (7).

² This acceptability metric may also be applied to cases in which the choice within a candidate set varies across instances of evaluation (see Boersma and Hayes 2001 on phonological variation and gradient acceptability). In HG theories of variation (Goldwater and Johnson 2003, Boersma and Pater 2008), a candidate that is chosen in a higher proportion of evaluations also has higher (mean) harmony. Thus, higher relative frequency entails higher acceptability.
We thus now have a three-way acceptability distinction: voiced codas are rated +1 by speakers of Voico-y, +0.1 by speakers of Voico-e, and –1 by speakers of Voico-n.

This approach can similarly capture the postulated difference between Voico-e’s marginal voiced obstruent codas, and its robustly attested palatalized onsets, as well as its absolute ban on palatalized codas. In the tableaux in (9), the faithful mapping /p'at/ → [p'at] wins by a relatively large margin because the difference between the weights of IDENT-PALATAL and *PALATAL is relatively large. A palatalized coda will receive a negative acceptability score so long as the summed weights of PALATAL → ONSET and *PALATAL are greater than that of IDENT-PALATAL, as shown in the second tableau.

To complete the analysis, I include a tableau for the marginal voiced codas that incorporates both *VOICE and the constraint against voiced codas, VOICE→ONSET. The weights in (11) give the voiced coda only a slim advantage over its competitor.

With these constraint weights, a voiced onset would get a positive score of 1.1, since it violates just *VOICE, and its competitor violates IDENT-VOICE. Depending on how sensitive speakers are to fine distinctions in the acceptability continuum, we now have a three- or four-way distinction for speakers of Voico-e: well-formed voiced obstruent (+1.1) and palatalized (+1) onsets, marginal voiced obstruent codas (+0.1), and ill-formed palatalized codas (–1).
As Coetzee and Pater (2008) and Hayes and Wilson (2008) point out, not only does HG offer a grammatical means of expressing gradient phonotactics, but it is also compatible with a range of learning algorithms developed in the context of statistical learning and neural modeling. To see if suitable constraint weights could be learned for our three hypothetical languages using the learning algorithm adopted by Coetzee and Pater (2008), I performed the following experiment using Praat (Boermsa and Weenink 2008).

The data for learning included three word types – ones with voiced obstruent onsets (‘[bat]’), ones with palatalized onsets (‘[p'at]’), and ones with voiced codas (‘[tad]’). The word types were presented to the learners in the following probability distributions.

\[
\begin{array}{c|ccc}
 & \text{Voico-y} & \text{Voico-n} & \text{Voico-e} \\
\hline
/bat/ \rightarrow [bat] & 0.33 & 0.50 & 0.498 \\
/p'at/ \rightarrow [p'at] & 0.33 & 0.50 & 0.498 \\
/tad/ \rightarrow [tad] & 0.33 & 0.00 & 0.005 \\
\end{array}
\]

The Voico-e distribution would be produced if a language had 100 words with voiced obstruent onsets, 100 words with palatalized onsets, and 1 word with a voiced obstruent coda, and these words were presented with equal frequency to the learner.\(^3\)

The learners were given one of two constraint sets. The “Positional Markedness” constraint set included all of the constraints introduced thus far. The “Positional Faithfulness” constraint set replaced \textit{VOICE}→\textit{ONSET} and \textit{PALATAL}→\textit{ONSET} with the following constraints:

\[
\begin{align*}
\text{(12) } & \text{ a. IDENT-VOICE-ONSET} \\
& \text{ Every pair of segments in correspondence in which the output segment is in onset position agrees in specification for [voice]} \\
& \text{ b. IDENT-PALATAL-ONSET} \\
& \text{ Every pair of segments in correspondence in which the output segment is in onset position agrees in specification for [palatal]} \\
\end{align*}
\]

The Positional Faithfulness constraint set was included because these constraints are known to exacerbate the phonotactic learning problem in OT (Hayes 2004, Prince and Tesar 2004, Tessier 2006). For an HG grammar to allow [voice] specifications in onset position, the sum of the weights of IDENT-VOICE→ONSET and IDENT-VOICE must be greater than that of *VOICE. Weights meeting this condition appear in (13), with the first tableau showing the faithful parsing of onset [voice]. For [voice] to be banned from coda position, the weight of IDENT-VOICE must be less than that of *VOICE. Because the weights in (13) also meet this condition, the optimal candidate in the second tableau undergoes coda devoicing.

\[\]  
\[\]

\(^3\) A more realistic learning procedure would specify the level at which phonotactic learning takes place: in the parsing of perceived utterances, in the construction of lexical entries, in lexical access, in production, or in some combination of all of these. It would also address the relative contribution of type and token frequency to phonotactic acceptability (see Albright 2007 for recent discussion).
Coda devoicing with positional faithfulness in HG

\[
\begin{array}{|c|c|c|c|}
\hline
\text{word type} & \text{*Voice} & \text{IDENT-VOICE-ONSET} & \text{IDENT-VOICE} \\
\hline
\text{/da/} & 1.5 & -1 & 1 \\
\text{[ta]} & -1 & & \\
\text{☞ [da]} & & & -2 \\
\hline
\text{/tad/} & 1.5 & -1 & 1 \\
\text{[tad]} & -1 & & \\
\text{☞ [tat]} & & & -1.5 \\
\hline
\end{array}
\]

For both constraint sets the markedness constraints were given an initial weight of 10, and the faithfulness constraints were given an initial weight of 0 (following Jesney and Tessier 2007).

The learners’ grammars used one of two evaluation methods. The HG learners used the method illustrated in the tableaux above: the outcome of evaluation is the optimal form in the candidate set. For the Maximum Entropy learners, the outcome of each evaluation is sampled from a probability distribution over the candidate set: the probability of a candidate is proportional to the exponential of its harmony (Goldwater and Johnson 2003).4

All learners used the same learning algorithm as in Coetzee and Pater (2008), termed HG-GLA in Boersma and Pater (2008), to which readers are directed for a thorough formal presentation. For the Maximum Entropy learners, this is stochastic gradient ascent, as in Jäger (to appear). HG-GLA is an on-line error driven learning algorithm in which the degree of change of a constraint’s weight is proportional to the difference in violation scores between the incorrect form that the learner’s grammar currently chooses and the correct form in the learning data. To the degree that a constraint is violated more in the correct form, its value is lowered, and to the degree that it is violated more in the incorrect form, it is raised. The learning rate, or plasticity was set at 0.1.

The results for the HG learners are presented in the following table.5 The numbers beneath each word type indicate the Acceptability score assigned to each one by the final state grammar.

---

4 This type of Maximum Entropy grammar is different from that in Hayes and Wilson (2008), in which the probability of word is calculated over the space of possible words.

5 I also ran these simulations with Noisy HG grammars (Boersma and Pater 2008). The result was similar to that in (15), except that the acceptability scores for attested forms were less close to zero, due to the need to overcome the noise.
These learners were successful in that every word that was in the dataset receives a positive acceptability score, and every one that was absent receives a negative score. This result held for both constraint sets, providing further support for Jesney and Tessier’s (2007) claim that HG learners have an advantage on phonotactic learning problems with positional faithfulness constraints. This advantage stems from the gang effect illustrated in (14): the learners stop making errors when the sum of the weights of the specific and general faithfulness constraints is greater than the weight of the markedness constraint, leaving the general faithfulness constraint beneath the markedness constraint. The final weights did not straightforwardly reflect relative frequency, however: [tad] receives an Acceptability score of +0.2 if there are voiced codas in the dataset, whether they are rare (Voico-e) or not (Voico-y). An effect of frequency is seen in that [da] received higher scores when voicing was more common in onset than when it was equally common in both positions. Thus, [tad] is less acceptable relative to [da] when voiced codas are rare. But this also affects the well-formedness of [da] relative to [p’at], which seems odd.

As shown in the following table, the Maximum Entropy learners were less successful in assigning negative scores to unseen forms: they always assigned positive scores to unattested forms when they were using the Positional Faithfulness constraints. Unlike the HG learners however, there is a more straightforward match to relative frequency: [tad] receives higher Acceptability scores from the Voico-y learners than the Voico-e ones.

I take these learning outcomes to be encouraging, in that they required only a very simple learning algorithm with a very simple markedness over faithfulness bias. Further research is required to better understand the differences between the models of grammar used by these learners, and to determine whether phonotactic learning requires different sorts of biases and/or explicit computation of expected distributions across the space of possible words, as claimed by Jarosz (2006) and Hayes and Wilson (2008).
3. Conclusions

In this paper I have used simple hypothetical examples of gradient phonotactics to highlight two attractive properties of an HG account. First, the acceptability of a form is calculated using the same harmony measure that determines optimality in the candidate set. As Coetzee and Pater (2008) point out, the comparison of harmony scores in their relativized acceptability measure is also independently motivated in HG-GLA and other learning algorithms. Second, HG-GLA and other learning algorithms produce final state grammars that yield acceptability scores that reflect the relative attestedness of different structures in the learning data.

Anttila’s (2008) OT-based Complexity Hypothesis has neither of these features. Anttila (2008: 2) defines it as follows:

(16) The hypothesis is that the relative well-formedness of a phonotactic structure depends on its grammatical complexity in the following sense: the more ranking information a phonotactic structure requires in order to surface faithfully, the less well-formed it is.

Anttila does not provide an explicit statement of how well-formedness is calculated in this approach, that is, how given a form, and a language-specific ranking, a degree of well-formedness is assigned. If one were to write an explicit algorithm for this calculation, it would likely be a fairly elaborate procedure, as is the calculation of gradient acceptability in Coetzee and Pater’s (2006) earlier OT account. Antilla (2008) also offers no alternative to the OT learning accounts proposed in Pater (2005), and adopted by Coetzee and Pater (2006), nor to the HG learning account in Coetzee and Pater (2008). Thus, the Occam’s razor argument that Antilla offers for the Complexity Hypothesis has little force, since the lexically specific constraints posited in Coetzee and Pater (2006) or the weighted constraints posited by Coetzee and Pater (2008) are independently required to get markedness constraints into the order that Anttila assumes in his discussion of Muna. OT’s Constraint Demotion algorithm (Tesar and Smolensky 2000) would not yield these rankings.

In sum, HG permits an account of how speakers compute gradient acceptability given a grammar, and also an account of how that grammar is learned, that appears to be more elegant, and more empirically adequate (see Coetzee and Pater 2008), than earlier OT-based accounts. The Complexity Hypothesis, on the other hand, does not appear to offer either of these advances.

References


---

6 Technically, HG-GLA subtracts the unweighted score on each constraint for the learner’s incorrect optimum from that of the correct learning datum to produce an error vector. The sum of the scores on that vector, each multiplied by its constraint’s weight, is equivalent to the acceptability of the incorrect form.


Coetzee, Andries W. and Joe Pater. 2006. ‘Lexically ranked OCP-Place constraints in Muna’, ms, University of Michigan and University of Massachusetts, Amherst (ROA-842).


Pater, Joe. 2008a. ‘Optimization and Linguistic Typology’. Ms, University of Massachusetts, Amherst.


