GRADIENT PHONOTACTICS IN OPTIMALITY THEORY

ABSTRACT: Lexical items can be more or less well-formed depending on the phoneme combinations they contain. This phenomenon is called gradient phonotactics. We propose that the relative well-formedness of a phoneme combination is inversely correlated with its grammatical complexity defined in terms of optimality-theoretic ranking information and show that this hypothesis is supported by phonotactic data from Muna (Austronesian). We conclude that gradient phonotactics does not require new theoretical devices, such as gradient constraints, but follows from the elementary notions of Optimality Theory: constraints, rankings, and relations among rankings.

1. INTRODUCTION

PHONOTACTICS is the study of permissible and impermissible phoneme combinations in a language. It has often been noted that phonotactic principles appear to be GRADIENT: lexical items can be more or less well-formed depending on the phoneme combinations they contain.

The gradience of phonotactics emerges in at least two ways. First, some types of lexical items are statistically overrepresented, others statistically underrepresented, depending on their phonotactic structure. For example, in Arabic, there is a well-known dissimilatory constraint against homorganic consonants in adjacent positions within the verbal root (Frisch, Pierrehumbert, and Broe 2004; Greenberg 1950; McCarthy 1988, 1994; Pater and Coetzee 2005; Pierrehumbert 1993): the more similar the consonants are, the less commonly they co-occur in actual lexical items. For similar gradient generalizations in other languages, see e.g. Berkley 1994a, 1994b, 2000; Coleman and Pierrehumbert 1997; Hammond 2004; Hay, Pierrehumbert, and Beckman 2004 (English); Coetzee and Pater 2006; Pater and Coetzee 2005 (Muna). Second, it has been observed that novel words ("wug words") show gradient acceptability that depends on their phonotactic structure. Thus, speakers of English judge nonsense words like *stin* to be rather good, *smy* to be less good, and *bzarshk* to be rather bad (Albright 2006). For similar effects, see Bailey and Hahn 2001; Coleman and Pierrehumbert 1997; Frisch, Large, and Pisoni 2000; Frisch and Zawaydeh 2001; Greenberg and Jenkins 1964; Ohala and Ohala 1986; Vitevitch, Luce, Charles-Luce, and Kemmerer 1997, among others.

Gradient phonotactic generalizations are a challenge for phonological theory. Nevertheless, they have attracted relatively little attention from phonologists. How should such generalizations be explained? There are two main possibilities. The first possibility is a GRAMMATICAL EXPLANATION: phonological grammars are formalized in such a way as to predict the relative likelihoods of segment combinations in terms of their relative markedness, perhaps stated over natural classes. Some segment combinations would be so ill-formed as to end up being absolutely ungrammatical, while others would be more or less grammatical along different dimensions of markedness. The second possibility is a LEXICAL EXPLANATION: gradient judgments arise by consulting the lexicon. On this view, a novel word would derive support from existing words depending on the number of its lexical neighbors, defined in terms of e.g. string edit distance (Kruskal 1983), which may be weighted by lexical token frequency, similarity, etc. An extreme version of such a model would deny the phonological grammar any role in gradient well-formedness judgments. The speakers would simply consult the existing lexical items in judging the well-formedness of novel words, not abstract markedness relations stated over combinations of natural classes.

It is likely that a successful explanation of gradient phonotactics will ultimately involve both grammatical and lexical factors. The best approach seems to be to develop explicit theories of both types and try to figure out what kind of division of labor is empirically justified. In this paper, we will pursue a grammatical explanation of gradient phonotactics based on Optimality Theory (Prince and Smolensky 1993/2004). Our goal is to derive the relative well-formedness of phonotactic combinations from a grammar based on ranked and violable constraints. Our proposal is that the relative well-formedness of a phonotactic combination depends on its grammatical complexity in the following sense: the more ranking information a phonological mapping requires, the less well-formed it is. We call this the COMPLEXITY HYPOTHESIS:

(1) The Complexity Hypothesis: The probability of an <input, output> mapping is inversely correlated with its grammatical complexity.

We show that the Complexity Hypothesis is consistent with the gradient phonotactic facts of Muna (Austronesian), recently studied by Coetzee and Pater (2006). The results support the view that gradient phonotactics is part of the phonological grammar, not just a by-product of lexical frequencies.

The present paper draws upon two lines of recent optimality-theoretic research. Our proposal is indebted to Pater and Coetzee's work on gradient phonotactics (Coetzee and Pater 2006, Pater and Coetzee 2005). We build on their insights, but develop them in a very different direction. While working on this project, I have become aware of Prince's work on Elementary Ranking Conditions (Prince 2002a, 2002b, 2006a, 2006b). This work seeks to determine the necessary and sufficient ranking conditions for phonological mappings and the entailments among these conditions. The present paper provides a concrete illustration of the relevance and usefulness of these notions in empirical work.

The paper is structured as follows. In section 2, we introduce our proposal and illustrate it based on a simple example from Arabic. In section 3, we discuss the gradient phonotactics of Muna. Section 4 concludes the paper.

2. THE PROPOSAL

2.1 *The Arabic example*

In Arabic, root morphemes are normally composed of three consonants, e.g. *ktb* 'write'. There is a well-known dissimilatory constraint against homorganic consonants in adjacent positions within the verbal root (Frisch, Pierrehumbert, and Broe 2004; Greenberg 1950; McCarthy 1988, 1994; Pater and Coetzee 2005; Pierrehumbert 1993). For example, root morphemes with adjacent labial consonants (**fbm, *bfk, *kbm*) are ill-formed (McCarthy 1988:88). However, it is also well known that the pattern shows gradience. Frisch, Pierrehumbert and Broe (2004) argue that the strength of the dissimilatory effect is a gradient function of the similarity of the consonants in the pair: the more similar the consonants, the less frequently they co-occur in actual lexical items.

Frisch, Pierrehumbert, and Broe (2004) describe the gradient phonotactic patterns of Arabic by means of OBSERVED/EXPECTED (O/E) VALUES. The O/E value is the ratio of the observed number of occurring consonant pairs (O) to the number that would be expected if the consonants combined at random (E). An O/E value greater than 1 indicates that there are more observed combinations than expected, i.e. the combination is favored. An O/E value smaller than 1 indicates that there are fewer observed combinations than expected, i.e. the combination is disfavored. The O/E values are shown in (2). Following Pater and Coetzee 2005, we have omitted dorsals and gutturals.

(2) Observed/Expected values for pairs of adjacent consonants in Arabic verbal roots (Frisch, Pierrehumbert, and Broe 2004:186):

	labial	dorsal	coronal sonorant	coronal fricative	coronal plosive
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

labials:	b,f,m
dorsals:	k, g, q
coronal sonorants:	l, r, n
coronal fricatives:	θ, ð, s, z, $s^ʕ$, $z^ʕ$, ʃ
coronal plosives:	t, d, $t^ʕ$, $d^ʕ$

The central observation can be stated as follows. The O/E values for pairs of homorganic consonants are near zero (shaded cells), but within coronals there is gradience (bottom right hand box):

(3) The quantitative patterning of adjacent coronals

(a) If the coronals are both sonorants, fricatives, or plosives, O/E is low;
(b) If the coronals are fricative + plosive, O/E is higher;
(c) If the coronals are sonorant + fricative or plosive, O/E is high.

The challenge is to derive this quantitative pattern from the grammar.

2.2 *T-orders*

Following McCarthy (1988, 1994) and Padgett (1995), Pater and Coetzee (henceforth P&C 2005) posit a set of OCP-constraints against identical feature specifications on adjacent segments where adjacency is defined on the consonantal root. Their constraints are listed in (4).

(4) Constraints (P&C 2005)

FAITH	Input and output are identical.
OCP-COR	No adjacent coronals (e.g. /t-n/)
OCP-COR[−son]	No adjacent coronal obstruents (e.g. /t-s/)
OCP-COR[+son]	No adjacent coronal sonorants (e.g. /l-n/)
OCP-COR[−son][αcont]	No adjacent coronal obstruents agreeing in continuancy (e.g. /t-d/)

Let us examine the constraint violations for four inputs: plosive-plosive (/t-d/), plosive-fricative (/t-s/), plosive-sonorant (/t-n/), and sonorant-sonorant (/l-n/). The violations are shown in (5). OTHER stands for the unfaithful candidate that violates FAITH. No rankings are intended.[1]

(5) The constraint violations for coronals

		FAITH	OCP-COR [−son, αcont]	OCP-COR [+son]	OCP-COR [−son]	OCP-COR
t-d	t-d		*		*	*
	OTHER	*				
t-s	t-s				*	*
	OTHER	*				
t-n	t-n					*
	OTHER	*				
l-n	l-n			*		*
	OTHER	*				

[1] Using all six coronal cells would not introduce any new types of constraint violation patterns.

The standard assumption in Optimality Theory is that constraints are universal, rankings language-particular. Given these five constraints, the question is what kinds of languages they predict to be possible and what kinds of languages they exclude as impossible. We can find out the answer by computing the FACTORIAL TYPOLOGY using OTSOFT ranking software (Hayes, Tesar, and Zuraw 2003). The software considers all the 120 total rankings of the 5 constraints and finds all the predicted output patterns. Only 7 distinct output patterns are found. The factorial typology is shown in (6).

(6) The factorial typology for coronals

```
            #1    #2    #3    #4    #5    #6    #7
/t-d/:      t-d   t-d   OTHER OTHER OTHER OTHER OTHER
/t-s/:      t-s   t-s   t-s   t-s   OTHER OTHER OTHER
/t-n/:      t-n   t-n   t-n   t-n   t-n   t-n   OTHER
/l-n/:      l-n   OTHER l-n   OTHER l-n   OTHER OTHER
```

The factorial typology in (6) reveals several asymmetries. For example, there are two languages where /t-d/ surfaces faithfully (#1, #2). These languages are a subset of the four languages where /t-s/ surfaces faithfully (#1, #2, #3, #4). This generalization can be stated as an IMPLICATIONAL UNIVERSAL in terms of <input, output> pairs as follows:

(7) An implicational universal

`<t-d, t-d> --> <t-s, t-s>` If /t-d/ is realized faithfully, so is /t-s/.

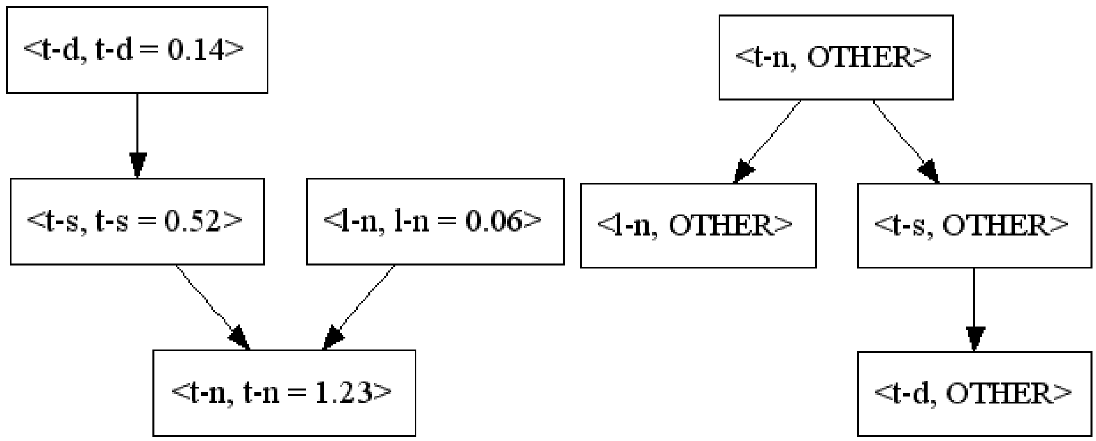
The implicational universal (7) is true of every language in the factorial typology. In other words, it is true no matter how the constraints are ranked. However, implicational universals need not be immediately obvious. Apart from the simplest cases, they are not easy to read off the constraint definitions ((4)), violation patterns ((5)), or factorial typologies ((6)). This raises the possibility that there are more implicational universals hidden in the factorial typology. If we systematically work through the factorial typology in (6), we discover eight implicational universals, summarized in (8). We call the set of all implicational universals in a factorial typology a T- ORDER.

(8) T-order as pairs of <input, output> pairs

```
(a)  <t-d, t-d> --> <t-s, t-s>
(b)  <t-d, t-d> --> <t-n, t-n>
(c)  <t-s, t-s> --> <t-n, t-n>
(d)  <l-n, l-n> --> <t-n, t-n>
(e)  <t-s, OTHER> --> <t-d, OTHER>
(f)  <t-n, OTHER> --> <t-d, OTHER>
(g)  <t-n, OTHER> --> <t-s, OTHER>
(h)  <t-n, OTHER> --> <l-n, OTHER>
```

The structure in (8) is explicit, but somewhat unintuitive. It becomes easier to understand if we visualize it as a directed graph. This is shown in (9). The pair notation in (8) is equivalent to the graph notation in (9). The nodes have been annotated by the Arabic O/E values taken from Frisch, Pierrehumbert, and Broe 2004.

(9) T-order as a directed graph



The graph in (9) reveals an interesting fact: O/E values grow in the direction of the arrows. This means that the implicational universals derived by the grammar emerge quantitatively in the Arabic lexicon. Why should this be?

The key observation is that the T-order reflects the grammatical complexity of <input, output> mappings. Consider again the violation patterns, repeated in (10):

(10) Violation patterns for coronals

		FAITH	OCP-COR [−son, αcont]	OCP-COR [+son]	OCP-COR [−son]	OCP-COR
t-d	t-d		*		*	*
	OTHER	*				
t-s	t-s				*	*
	OTHER	*				
t-n	t-n					*
	OTHER	*				
l-n	l-n			*		*
	OTHER	*				

What rankings are required for the faithful candidate to win? The answer is that FAITH must dominate every constraint violated by the faithful candidate. For example, the mapping </t-d/, [t-d]> requires the conjunction of three pairwise rankings to win:

(11) The rankings required for the faithful mapping </t,d/, [t-d]>

FAITH >> OCP-COR ∧
FAITH >> OCP-COR[−son] ∧
FAITH >> OCP-COR[−son, αcont]

The graph in (12) shows all the faithful mappings and their required rankings ordered by the amount of ranking information. Each step downwards removes a conjunct from the ranking statement, making it less informative. We have rediscovered the left hand side of the T-order in (9).

(12) Faithful mappings ordered by the amount of ranking information

</t,d/, [t-d]> (O/E = 0.14):
FAITH >> OCP-COR ∧
FAITH >> OCP-COR[−son] ∧
FAITH >> OCP-COR[−son, αcont]

↓

</t,s/, [t-s]> (O/E = 0.52):
FAITH >> OCP-COR ∧
FAITH >> OCP-COR[−son]

</l,n/, [l-n]> (O/E = 0.06)
FAITH >> OCP-COR ∧
FAITH >> OCP-COR[+son]

↓

</t,n/, [t-n]> (O/E = 1.23)
FAITH >> OCP-COR

What rankings are required for the unfaithful candidate (= OTHER) to win? The answer is that FAITH must be dominated by at least one constraint violated by the faithful candidate. For example, the mapping </t-n/, OTHER> requires one pairwise ranking to win:

(13) The rankings required for the unfaithful mapping </t-n/, OTHER>

OCP-COR >> FAITH

The graph in (14) shows all the unfaithful mappings and their required rankings ordered by the amount of ranking information. Each step downwards adds a disjunct to the ranking statement, making it less informative. We have rediscovered the right hand side of the T-order in (9).

(14) Unfaithful mappings ordered by the amount of ranking information

</t-n/, OTHER>:
OCP-COR >> FAITH

</l-n/, OTHER>:
OCP-COR >> FAITH ∨
OCP-COR[+son] >> FAITH

</t-s/, OTHER>:
OCP-COR >> FAITH ∨
OCP-COR[−son] >> FAITH

</t,d/, OTHER>:
OCP-COR >> FAITH ∨
OCP-COR[−son] >> FAITH ∨
OCP-COR[−son, αcont] >> FAITH

We have now arrived at two different ways of looking at T-orders: EXTENSIONALLY in terms of factorial typologies and INTENSIONALLY in terms of rankings. We started by defining T-orders in terms of implications among <input, output> mappings in the factorial typology. We then observed that the T-order arranges the <input, output> mappings by decreasing complexity, where complexity is measured in terms of the amount of ranking information. We can now state the quantitative generalization in grammatical terms as follows:

(15) The Complexity Hypothesis (preliminary version): The O/E value of a phonotactic sequence is inversely correlated with its grammatical complexity.

It is commonly assumed that a constraint ranking is a property of an entire language. Under this view, there would be one single ranking for Arabic. This ranking would have to include the pairwise rankings required for the mapping </t,d/, [t-d]> since this mapping is possible (albeit dispreferred) in the language. This ranking is shown in (16).

(16) The ranking for Arabic (first attempt)

		FAITH	OCP-COR [−son, αcont]	OCP-COR [+son]	OCP-COR [−son]	OCP-COR
t-d	☞ t-d		*		*	*
	OTHER	*				
t-s	☞ t-s				*	*
	OTHER	*				
t-n	☞ t-n					*
	OTHER	*				
l-n	☞ l-n			*		*
	OTHER	*				

The ranking in (16) leaves us with a puzzle. A central assumption in Optimality Theory is that the number of constraint violations plays no role in determining the grammaticality of a candidate as long as the candidate is optimal. There is thus no sense in which one faithful mapping in (16) should be worse than another. We thus have no explanation for the observed ordering of O/E values.

An alternative view holds that rankings are properties of lexical items. This is the COPHONOLOGY THEORY (see e.g. Anttila 2002, 2006; Inkelas 1998; Inkelas and Zoll 2003; Orgun 1996; Raffelsiefen 1999; Zamma 2005). Under this theory, the ordering among the O/E values is explained if we assume that the learner values simplicity in the lexicon. A lexical item that requires more ranking information is at a disadvantage in comparison to a lexical item that requires less ranking information. This is reflected quantitatively in the Arabic lexicon: lexical items that require more ranking information are less frequent than lexical items that require less ranking information.[2]

2.3 *Computing T-orders*

Given an optimality-theoretic grammar, how can we find the T-order? The Arabic example suggested two possibilities: an indirect method based on factorial typologies and a direct method based on constraint violation patterns.

The indirect method involves two steps. The first step is to compute the factorial typology. The second step is to retrieve the T-order from the factorial typology by checking which <input, output> pairs universally imply which other <input, output> pairs. In our Arabic example, this was relatively easy to do because the factorial typology was small. However, in general, factorial typologies are hard for humans to understand. OTSOFT can easily compute very large factorial typologies, but the result is usually too complex to make immediate sense. Even if the factorial typology is small, figuring out

[2] Under the Cophonology Theory, it is possible to talk about the ranking for a language, but only in a derivative sense: the ranking for a language is the intersection of the rankings of its lexical items. This partial ranking is called the MASTER RANKING by Inkelas and Zoll (2003).

the T-order with paper and pencil can be a tedious exercise. In analyses of realistic size, the number of output patterns may well run in the hundreds.

The direct method sidesteps factorial typologies by extracting T-orders directly from constraint violation patterns. This involves finding the required rankings for each <input, output> pair and determining the entailments among these rankings. This approach has been developed in several articles by Prince (2002a, 2002b, 2006a, 2006b) who formalizes the required rankings in terms of ELEMENTARY RANKING CONDITIONS (ERCs) and provides a calculus for working with ERCs. Rather than attempting a summary of Prince's work here, we refer the reader to the original papers. In our Arabic example, the entailments among the rankings were easy to find because we only considered two output candidates for each input. In general, finding the entailments is non-trivial and we face similar obstacles as with the indirect method.

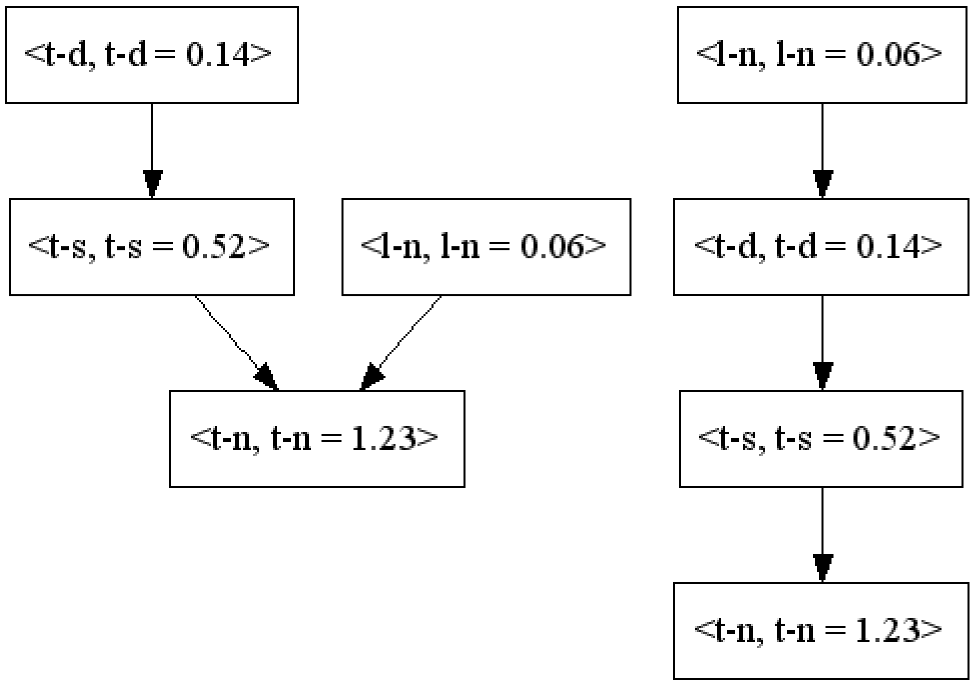
As noted in Prince (2006a), the two approaches to T-order computation represent two ways of looking at the same problem. Both approaches have the same goal: to discover the implications among the <input, output> pairs. Whichever approach we choose, the only viable option is to use a computer. All the T-orders in this paper have been computed by T-ORDER GENERATOR (Anttila and Andrus 2006), a Windows program that computes T-orders and visualizes them as directed graphs. The current version of the program (April 2007) allows the user to compute T-orders either indirectly from factorial typologies or directly from constraint violation patterns. The program reads OTSoft factorial typologies (indirect method) or Excel files written in the traditional tableau format (direct method). The direct method is implemented using ERCs which remain invisible to the user. The program is freely available from the present author's web site.

2.4 *Evaluating T-orders*

It is useful to have some way of determining how well a T-order matches the data. One possibility is to compare the actual T-order derived by the grammar to an ideal T-order that captures all the quantitative relationships in the data. The actual and ideal T-orders for the Arabic example are shown below in two different formats: (17) uses the graph notation; (18) uses the pair notation. The unfaithful half of the T-order is omitted because we only have O/E values for the faithful half.

(17) Actual and ideal T-orders for Arabic as directed graphs

(a) The actual T-order (b) The ideal T-order



(18) Actual and ideal T-orders for Arabic as pairs of pairs

```
(a)   <t-d, t-d> --> <t-s, t-s>          actual, ideal
(b)   <t-d, t-d> --> <t-n, t-n>          actual, ideal
(c)   <t-s, t-s> --> <t-n, t-n>          actual, ideal
(d)   <l-n, l-n> --> <t-n, t-n>          actual, ideal
(e)   <l-n, l-n> --> <t-d, t-d>                  ideal
(f)   <l-n, l-n> --> <t-s, t-s>                  ideal
```

In this case, the actual and ideal T-orders are not identical. In order to determine how well the two match, we calculate two evaluation measures commonly used in information retrieval: PRECISION and RECALL (see e.g. Manning and Schütze 1999: 267-271).

(19) Measuring the goodness of a T-order

(a) PRECISION is the ratio of the number of pairs that are in both T-orders to the number of pairs that are in the actual T-order. This number indicates how many of the predicted quantitative relationships are observed. In our example, precision = 4/4 = 100%.

(b) RECALL is the ratio of the number of pairs that are in both T-orders to the number of pairs that are in the ideal T-order. This number indicates how many of the observed quantitative relationships are predicted. In this example, recall = 4/6 = 67%.

What do these numbers mean? Assume that the T-order is based on universal constraints, but no language-specific rankings. Perfect precision (= 100%) means that all the predicted quantitative relationships are observed in the data. In other words, the universals have stood the test of one particular language. In this sense, high precision is an indicator of DESCRIPTIVE SUCCESS. Perfect recall (= 100%) means that all the quantitative relationships observed in the data are predicted. In other words, they are all universal. Given the possibility of language-specific variation, this seems an unlikely event. However, the higher the recall value, the fewer language-specific stipulations are needed to account for the residual facts. In our example, 67% of the quantitative relationships are universal; the rest have to be learned from the Arabic data. In this sense, high recall is an indicator of EXPLANATORY SUCCESS.

2.5 *T-orders and rankings*

In the Arabic example, the T-order was computed with no rankings. One may wonder how adding rankings would change the shape of the T-order. It turns out that adding rankings can add arrows into the T-order, but never subtract any existing arrows. In order to see this, consider the possible relations between <input, output> mappings in the factorial typology. There are three distinct cases:

(20) Three types of relations between <input, output> mappings

(a) Neither mapping implies the other (= no implication):

```
             #1    #2    #3    #4    #5    #6    #7
/t-s/:       t-s   t-s   t-s   t-s   OTHER OTHER OTHER
/l-n/:       l-n   OTHER l-n   OTHER l-n   OTHER OTHER
```

(b) One mapping implies the other (= implication):

```
             #1    #2    #3    #4    #5    #6    #7
/t-d/:       t-d   t-d   OTHER OTHER OTHER OTHER OTHER
/t-s/:       t-s   t-s   t-s   t-s   OTHER OTHER OTHER
```

(c) Both mappings imply the other (= equivalence):

```
             #1    #2    #3    #4    #5    #6    #7
/t-n/:       t-n   t-n   t-n   t-n   t-n   t-n   OTHER
/s-n/:       s-n   s-n   s-n   s-n   s-n   s-n   OTHER
```

Adding rankings into the grammar will either eliminate columns or have no effect. Clearly, adding rankings can never add new columns into the factorial typology. Now, consider the possible effects of ranking in each case. In (20a), we start out with no arrows. Eliminating #5 would add an arrow between [l-n] and [t-s]; eliminating #2 and #4 would add an arrow between [t-s] and [l-n]. In (20b), we start out with an arrow between [t-d] and [t-s]. Removing #3 and #4 would add an arrow between [t-s] and [t-d], resulting in arrows both ways, i.e. equivalence, but it is not possible to remove the existing arrow. This is because the implicational universal holds true of every column in the factorial typology and we can only remove columns. Finally, in (20c) we start out with arrows both ways, i.e. equivalence, and neither adding nor subtracting arrows is possible. We conclude that adding rankings can only add arrows into the T-order, but never remove any existing arrows.

2.6 *Summary*

T-orders are a consequence of standard Optimality Theory, not a new theoretical device. For this reason, they need no independent motivation or justification. An entirely different question is how T-orders should be interpreted empirically. We have proposed one such interpretation, the Complexity Hypothesis, which is an empirical claim that can be true or false. This can only be determined by detailed analyses of particular languages.

Given a set of constraints, the present theory makes specific and often surprising predictions about universal and language-particular phonotactic patterns. Given P&C 2005's constraints for Arabic, the theory predicts that the relative ordering of O/E values among [t-d] ≤ [t-s] ≤ [t-n] and [l-n] ≤ [t-n] should be universal, whereas the relative ordering of O/E values between [l-n] and [t-s] and between [l-n] and [t-d] should be able to vary from language to language. If we further assume that constraints are innate, it follows that a newborn knows all this before hearing the first word. Only the ranking-dependent facts must be learned from the data.

We introduced T-orders in the context of gradient phonotactic patterns. However, T-orders are in no way limited to phonotactics. A T-order is simply a relation (partial order) that holds among <input, output> mappings. The nature of the mapping is immaterial. In the case of phonotactics, the mapping is faithful, e.g. </t-d/, [t-d]> (no change). In the case of alternations, the mapping is unfaithful, e.g. </cost me/, [cos me]> (*t*-deletion). This means that T-orders and the Complexity Hypothesis generalize beyond phonotactics. Relevant studies include Anttila 2007a, Anttila and Andrus 2006, Anttila, Fong, Benus, and Nycz in progress (phonological alternations); Anttila 2007b (syntactic alternations); Blumenfeld 2005 (metrics); van Oostendorp and Hinskens 2007 (segmental inventories). A more general version of the Complexity Hypothesis that is neutral with respect to the nature of the <input, output> mapping is stated in (21):

(21) <u>The Complexity Hypothesis (final version)</u>: The probability of an <input, output> mapping is inversely correlated with its grammatical complexity.

3. PATER AND COETZEE 2005

Our proposal draws upon earlier work on gradient phonotactics by Pater and Coetzee (2005) (= P&C 2005) and Coetzee and Pater (2006). The proposals are similar in several respects. Both derive gradient phonotactics from the grammar; both assume ranked and violable constraints; and both aim at accounting for the relative well-formedness of lexical items. In order to see the differences, a brief review of Pater and Coetzee's theory is in order.

Recall the phonotactic patterns among Arabic coronals:

(22) The quantitative patterning of adjacent coronals

(a) If the coronals are both sonorants, fricatives, or plosives, O/E is low;
(b) If the coronals are fricative + plosive, O/E is higher;
(c) If the coronals are sonorant + fricative or plosive, O/E is high.

P&C 2005's analysis is built around the idea that markedness (here: OCP) constraints are ranked in an order that reflects the relative O/E values of the respective structures (here: consonant pairs). They posit the following ranking that reflects the Arabic O/E values:

(23) The ranking among OCP-COR constraints in Arabic (P&C 2005:6)

{OCP-COR[−son][αcont] OCP-COR[+son]}	>> OCP-COR[−son]	>> OCP-COR
violated by (22a), low O/E	violated by (22b), intermediate O/E	violated by (22c), high O/E

Gradience is captured in this model by positing multiple faithfulness constraints interspersed among the OCP constraints and by assuming that lexical items are randomly coindexed with these constraints. Each lexical item is assigned a well-formedness score by submitting it to the grammar with each lexical indexation. This is illustrated for the input /t-d/ in (24).

(24) Evaluating /t-d/ (Pater and Coetzee 2005)

		FAITH-L2	OCP-COR [−son, αcont]	OCP-COR [+son]	**FAITH-L1**	OCP-COR [−son]	**FAITH**	OCP-COR
(a) t-d	☞t-d		*			*		*
	OTHER	*!						
(b) t-d	t-d		*!			*		*
	☞OTHER				*			
(c) t-d	t-d		*!			*		*
	☞OTHER						*	

The input /t-d/ is indexed to FAITH-L2 in (24a), FAITH-L1 in (24b), and FAITH in (24c). The faithful candidate [t-d] only wins in (24a), i.e. under 1/3 of the indexations. This well-formedness score is interpreted relatively: the higher the score, the more acceptable the lexical item. The well-formedness scores for other inputs are derived in an analogous manner. The results of evaluating the four sample inputs are shown in (25).

(25) Well-formedness scores for four faithful mappings (P&C 2005)

	MAPPING	SCORE	O/E
(a)	/l-n/ → [l-n]	1/3	0.06
(b)	/t-d/ → [t-d]	1/3	0.14
(c)	/t-s/ → [t-s]	2/3	0.52
(d)	/t-n/ → [t-n]	3/3	1.23

In sum, P&C 2005's theory has three parts:

(26) An outline of P&C 2005's theory
- (a) Markedness constraints are ranked according to the frequency with which they are violated in the lexicon
- (b) Interspersed among the markedness constraints are lexically-specific faithfulness constraints.
- (c) The well-formedness of a lexical item is determined by submitting it to the grammar with each lexical indexation.

Our proposal is different in all three respects. We started by observing that phonological mappings can be ordered in terms of their grammatical complexity and used the resulting structure (T-order) to derive the ordering of O/E values. This is the key difference from which all other differences follow. First, we proposed that the well-formedness of a lexical item is inversely correlated with its grammatical complexity (The Complexity Hypothesis) (cf. (23c)). Second, we noted that some aspects of gradient phonotactics are ranking-dependent, hence language-specific, whereas other aspects are ranking-independent, hence universal. In the Arabic example, 67% of the quantitative relationships turned out to be independent of ranking (cf. (24a)). Finally, lexically-specific faithfulness constraints are not needed at all (cf. (23b)).

4. GRADIENT PHONOTACTICS IN MUNA

We now turn to the gradient phonotactics of the Austronesian language Muna (van den Berg 1989, van den Berg and Sidu 1996). Our discussion draws upon the detailed analysis in Coetzee and Pater 2006 (henceforth C&P 2006). Coetzee and Pater describe the data in great detail, down to the level of individual segment pairs, providing a good testing ground for theories of gradient phonotactics.

4.1 *The phonotactics of obstruents and nasals*

C&P 2006 first account for the phonotactics of Muna obstruents and nasals. Liquids are brought into the discussion later. We follow this presentational order here.[3] The analysis focuses on those segment types that can be compared across places of articulation. These segments are listed in (27).

[3] The phonotactics of prenasals and identical segments are not analyzed by Coetzee and Pater (2006).

(27) Muna obstruents and nasals:

labials	/p, b, f, m/
coronals	/t, d, s, n/
dorsals	/k, g, ʁ, ŋ/

The following segment pairs are compared:

(28) Muna obstruents and nasals:

	LAB	COR	DOR
voiced stop + voiceless stop	b-p	d-t	g-k
nasal + voiced stop	m-b	n-d	ŋ-g
nasal + voiceless stop	m-p	n-t	ŋ-k
nasal + fricative	m-f	n-s	ŋ-ʁ
fricative + voiced stop	f-b	s-d	ʁ-g
fricative + voiceless stop	f-p	s-t	ʁ-k

C&P 2006 start by positing the OCP constraints in (29). Note that coronals do not have the [αson, βcont] version of the OCP. An explanation is given in C&P 2006, p. 21.

(29) Constraints for Muna obstruents and nasals

OCP-DOR[αson, βcont]	OCP-LAB[αson, βcont]	--
OCP-DOR[αvoice]	OCP-LAB[αvoice]	OCP-COR[αvoice]
OCP-DOR[αcont]	OCP-LAB[αcont]	OCP-COR[αcont]
OCP-DOR[αson]	OCP-LAB[αson]	OCP-COR[αson]
OCP-DOR	OCP-LAB	OCP-COR

We assume that a general OCP-constraint (e.g. OCP-LAB) is violated iff the two adjacent consonants share place of articulation. A specific OCP-constraint (e.g. OCP-LAB[αcont]) is violated iff the two adjacent consonants share place of articulation and the value of the specified feature. This is illustrated in (30).

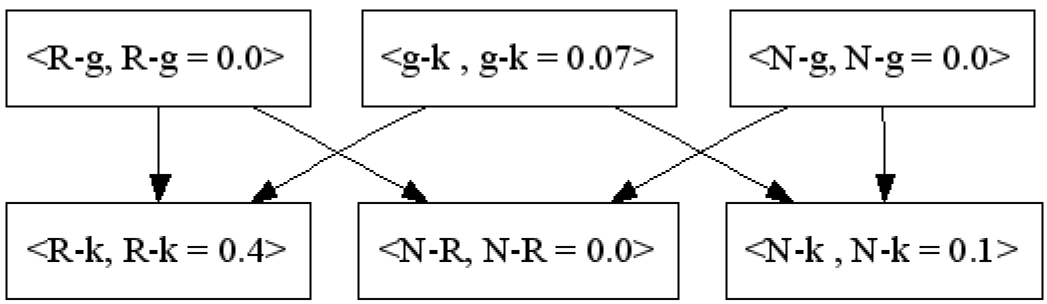
(30) The constraint violations for /m-p/, /f-b/ and /m-f/

	FAITH	OCP-LAB [αcont]	OCP-LAB [αson]	OCP-LAB
(a) m-p		*		*
OTHER	*			
(b) f-b			*	*
OTHER	*			
(c) m-f				*
OTHER	*			

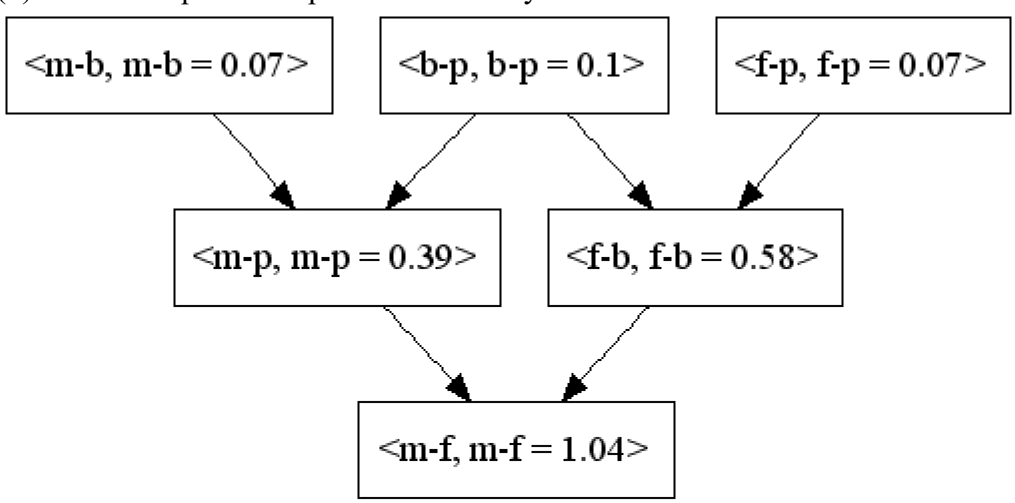
We start by computing the T-order based on the eighteen inputs in (28) and the fourteen constraints in (29), without assuming any rankings. The resulting T-order consists of three disjoint graphs, one for each place of articulation. These graphs are shown in (31). The nodes have been annotated by the O/E values from C&P 2006 who base themselves on 5,854 relevant roots from an electronic version of van den Berg and Sidu's (1996) Muna dictionary. Notice two typographical substitutions: N = ŋ, R = ʁ.

(31) T-order with no rankings: precision = 0.958, recall = 0.144

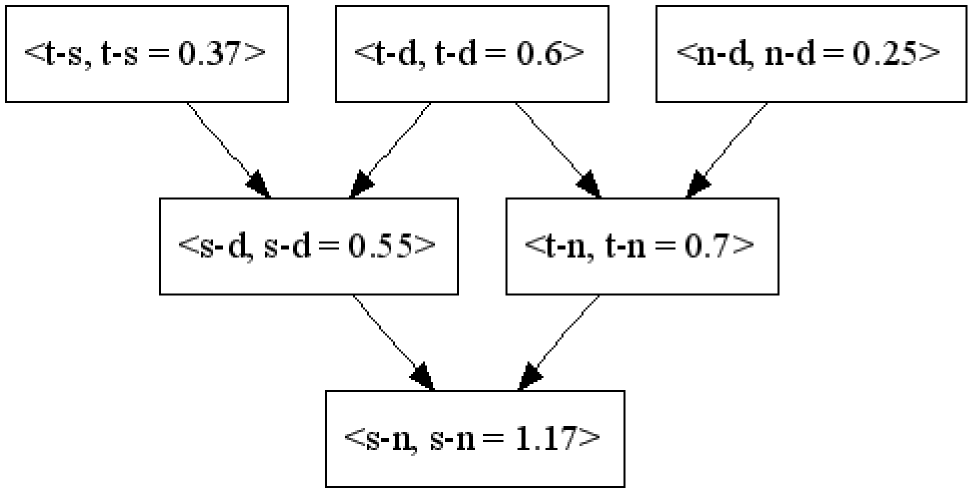
(a) Dorsal pairs. All pairs are correctly ordered.



(b) Labial pairs. All pairs are correctly ordered.



(c) Coronal pairs. All pairs except one ([t-d] ≤ [s-d]) are correctly ordered.



The predicted implicational universals hold up in the dictionary very well. The precision value (precision = 0.958) suffers from one minor error: [t-d] (O/E = 0.60) and [s-d] (O/E = 0.55) come out in the wrong order. The low recall value (recall = 0.144) indicates that there are many quantitative relationships in the data that are not captured directly by the universal constraints. This suggests that language-specific rankings are needed as well.

Why is [s-d] incorrectly predicted to have a higher O/E value than [t-d]? The violation pattern in (32) diagnoses the problem. The two mappings share the same violations except that [t-d] also violates OCP-COR[αcont]. This requires the ranking FAITH >> OCP-COR[αcont]. Since [s-d] requires fewer rankings than [t-d], the former sinks below the latter in the T-order.

(32) Comparing the violation patterns of </s-d/, [s-d]> and </t-d/, [t-d]>

		FAITH	OCP-COR[αcont]	OCP-COR[αson]	OCP-COR
s-d	s-d (O/E = 0.55)			*	*
	OTHER	*			
t-d	t-d (O/E = 0.60)		*	*	*
	OTHER	*			

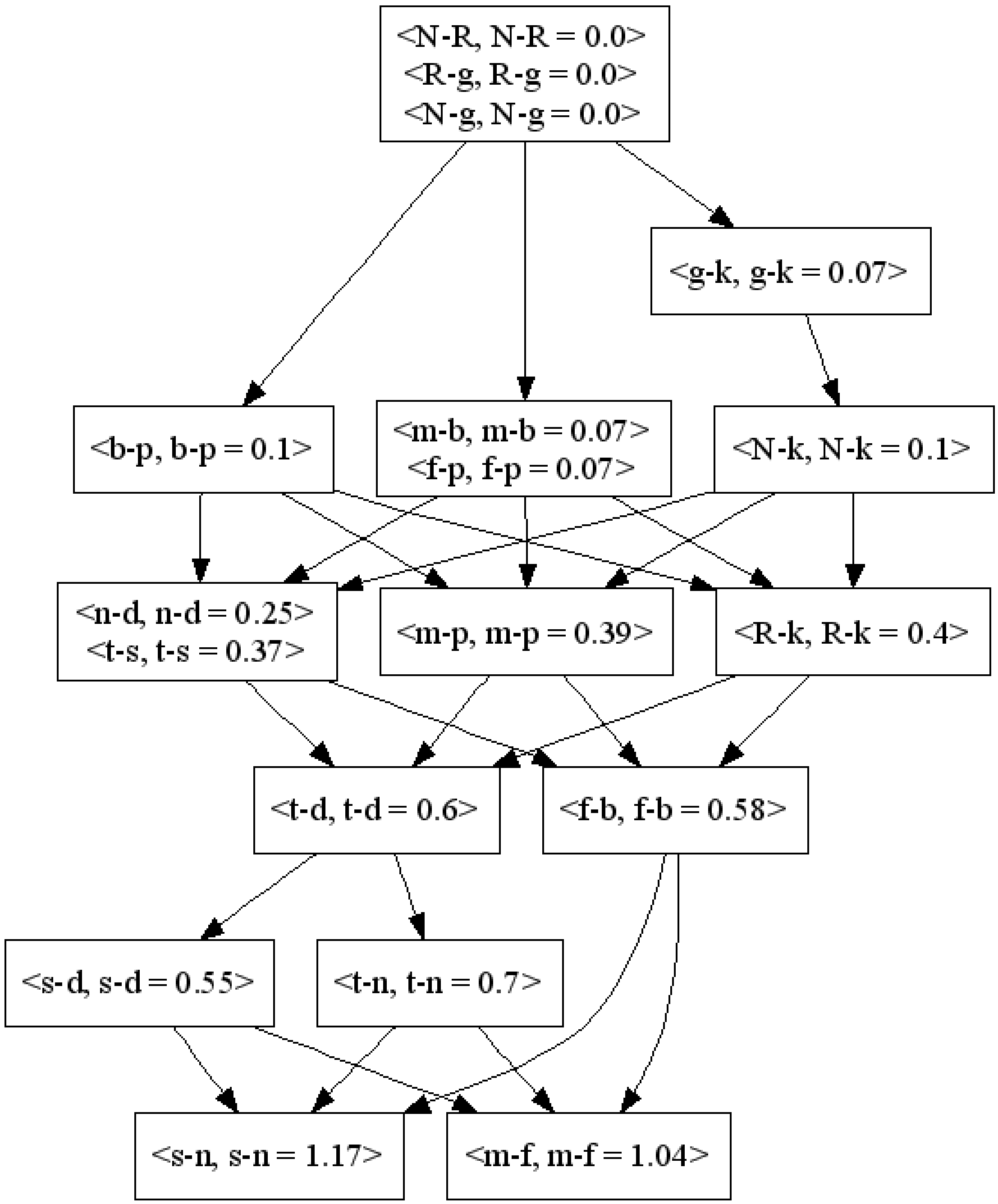
In addition to the constraints in (29), C&P 2006 posit rankings specific to Muna. They establish five internally unranked strata using the following heuristic: the smaller the O/E value of a consonant pair, the higher-ranked the markedness constraint targeting that pair. The ranking for Muna is shown in (33). The five strata correspond to the five groups of O/E values listed in the right hand column. The O/E values range from unattested (e.g. [ŋ-g], O/E = 0) to overrepresented (e.g. [n-s], O/E = 1.17).

(33) The constraint ranking for Muna

	TARGETED PAIRS	O/E VALUES
Stratum 1:		
OCP-DOR[αvoice]	ŋ-g, ʁ-g, ŋ-ʁ	0, 0, 0
Stratum 2:		
OCP-DOR[αson, βcont]	g-k	0.07
OCP-DOR[αcont]	ŋ-k	0.10
OCP-LAB[αson, βcont]	b-p	0.10
OCP-LAB[αvoice]	f-p, m-b	0.07, 0.07
Stratum 3:		
OCP-DOR[αson]	ʁ-k	0.40
OCP-DOR		
OCP-LAB[αcont]	m-p	0.39
OCP-COR[αvoice]	n-d, s-t	0.25, 0.37
Stratum 4:		
OCP-LAB[αson]	f-b	0.58
OCP-COR[αson]	s-d, d-t	0.55, 0.60
OCP-COR[αcont]	d-t, n-t	0.60, 0.70
Stratum 5:		
OCP-LAB	m-f	1.04
OCP-COR	n-s	1.17

We now add these rankings into the grammar and recompute the T-order. The result is shown in (34).

(34) The T-order for Muna obstruents and nasals based on C&P 2006's constraints and rankings: precision = 0.986, recall = 0.863



The graph in (34) shows that C&P 2006's constraints and rankings predict the ordering of the O/E values very well. The precision value is now nearly perfect (precision = 0.986) and the recall value improves dramatically (recall = 0.863). There are two minor problems on the precision side. The first problem is familiar: [t-d] (O/E = 0.60) and [s-d] (O/E = 0.55) come out in the wrong order. This problem is inherited from the constraints and for this reason cannot be fixed by adding rankings. The second problem originates from the rankings: [t-s] (O/E = 0.37) and [n-d] (O/E = 0.25) are predicted to have identical O/E values, but in reality there is a difference. The problem lies in the high-ranked OCP-COR[αvoice]. The grammar has two constraints that distinguish between [t-s] and [n-d], namely OCP-COR[αcont] and OCP-COR[αson], but the higher-ranked OCP-COR[αvoice] makes them irrelevant.

(35) OCP-COR[αvoice] is ranked too high

		FAITH	OCP-COR [αvoice]	OCP-COR [αcont]	OCP-COR [αson]	OCP-COR
n-d	n-d (O/E = 0.25)		*	*		*
	OTHER	*				
t-s	t-s (O/E = 0.37)		*		*	*
	OTHER	*				

These minor problems aside, the coverage of the analysis is very good. Under the present theory, C&P 2006's constraints and rankings for obstruents and nasals turn out 99% correct (precision = 0.986) and account for 86% of the quantitative relationships among the descriptive categories in their data (recall = 0.863). The residual patterns may well yield to a more fine-grained grammatical analysis or have an extragrammatical explanation. The Muna evidence thus lends strong support to the view that gradient phonotactics largely derives from grammatical constraints, instead of just being a by-product of lexical frequencies.

We conclude this section on a note on the feature [voice]. One of C&P 2006's central claims is that [voice] plays an unusually central role in Muna phonotactics. This is important in the context of their argument because it appears to provide evidence for language-specific constraint ranking and hence ultimately for Optimality Theory:

> The most striking thing about this analysis is that across places of articulation, the constraint relativized to [voice] is highest ranked. (C&P 2006, p. 14)
>
> The differences between Muna and Arabic show that the individual manners are not in a fixed relationship; Muna clearly has higher-ranked OCP-PLACE[αvoice] constraints than does Arabic. (C&P 2006, p. 15)

What is the evidence for the high ranking of OCP-PLACE[αvoice]? Recall that C&P 2006 rank their constraints using the following heuristic: the smaller the O/E value of a consonant pair, the higher-ranked the markedness constraint targeting that pair. As shown in (36), the Muna O/E values are indeed consistent with the high ranking of OCP-PLACE[αvoice]:

(36) Ranking OCP-PLACE[αvoice] (C&P 2006)

OCP-DOR[αvoice]	>>	OCP-DOR[αson], OCP-DOR		
ŋ-g, ʁ-g, ŋ-ʁ (0, 0, 0)		ʁ-k (0.40)		
OCP-LAB[αvoice]	>>	OCP-LAB[αson]	>>	OCP-LAB
f-p, m-b (0.07, 0.07)		f-b (0.58)		m-f (1.04)
OCP-COR[αvoice]	>>	OCP-COR[αson]	>>	OCP-COR
n-d, s-t (0.25, 0.37)		s-d, d-t (0.55, 0.60)		n-s (1.17)

However, the T-order in (31) reveals that several of the orderings in (36) follow independently of rankings. These universal orderings are listed in (37). The C&P 2006 constraints thus predict that the same quantitative relationships should be found in Arabic.

(37) Some universals of OCP-PLACE[voice]

ʁ-g (0)	≤	ʁ-k (0.40)		
f-p (0.07)	≤	f-b (0.58)	≤	m-f (1.04)
m-b (0.07)	≤	m-p (0.39)	≤	m-f (1.04)
n-d (0.25)	≤	n-t (0.70)	≤	n-s (1.17)
s-t (0.37)	≤	s-d (0.55)	≤	n-s (1.17)

The high ranking of OCP-PLACE[voice] thus appears less important than C&P 2006 suggest. This is not just a matter of redundancy in the description. As we will see shortly, the high ranking of OCP-COR[αvoice] results in a number of mispredictions in the phonotactics of liquids.

4.2 *Adding on the liquids*

After having set up the basic analysis, C&P 2006 extend it to pairs that contain the liquids /l,r/, both of which are coronal. This results in 9 additional pairs:

(38) Pairs containing liquids

	COR
voiceless stop + /l/	t-l
voiceless fricative + /l/	s-l
voiced stop + /l/	d-l
voiceless stop + /r/	t-r
voiceless fricative + /r/	s-r
voiced stop + /r/	d-r
/l/ + /r/	l-r
nasal + /l/	n-l
/r/ + nasal	r-n

C&P 2006 introduce two additional constraints: OCP-COR[αcont, βvoice] which targets [n-d] and [l-r] and OCP-COR[αson, βvoice] which targets [t-s], [l-r], [n-l], and [r-n]. Note that both are [voice]-related constraints.

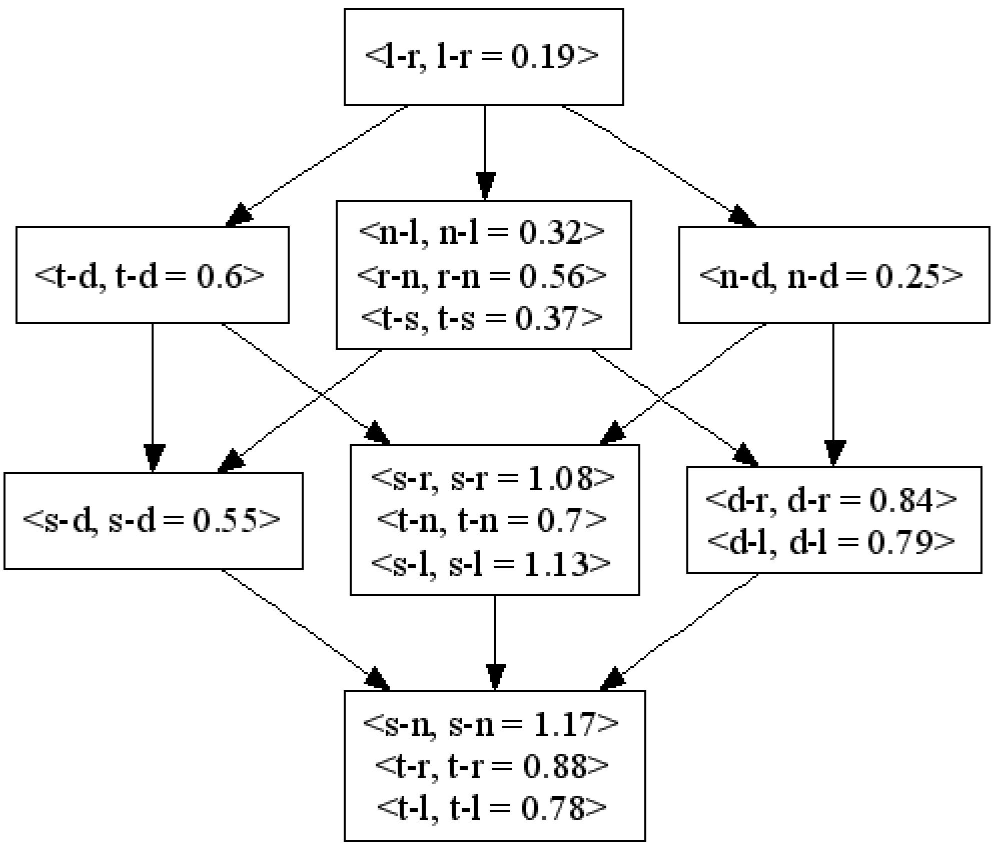
We start by recomputing the T-order, again first without rankings. The result is shown in (39). Only the coronal graph is shown; the labial and dorsal graphs remain the same as before.

(39) T-order with no rankings: precision = 0.820, recall = 0.229

(a) Dorsal pairs. All pairs are correctly ordered. (= (31))

(b) Labial pairs. All pairs are correctly ordered. (= (31))

(c) Coronal pairs:



We now compare the success of the augmented grammar (obstruents, nasals, liquids) to the success of the earlier grammar (obstruents, nasals). The grammar with liquids has a higher recall, but a lower precision than the grammar without liquids.

(40) Precision and recall before and after liquids, with no rankings

(a)	Obstruents, nasals:	Precision = 0.958	Recall = 0.144
(b)	Obstruents, nasals, liquids:	Precision = 0.820	Recall = 0.229

Why do we get the drop in precision? Three constraints turn out problematic: OCP-COR[αcont], OCP-COR[αson,βvoice] and OCP-COR[αvoice]. In what follows, we list all the precision errors. Note that most of them are numerically very small.

(41) The constraints responsible for the precision errors

(a) [s-r], [s-l] (O/E = 1.08, 1.13) ≤ [t-r], [t-l] (O/E = 0.88, 0.78). The problem arises because [s-r], [s-l] violate OCP-COR[αcont] whereas [t-r], [t-l] do not. This yields the wrong quantitative bias.

(b) [d-r], [d-l] (O/E = 0.84, 0.79) ≤ [t-l] (O/E = 0.78). The problem arises because [d-r], [d-l] violate OCP-COR[αvoice] whereas [t-l] does not. This yields the wrong quantitative bias.

(c) [r-n] (O/E = 0.56) ≤ [s-d] (O/E = 0.55). The problem arises because [r-n] violates OCP-COR[αson,βvoice] and OCP-COR[αvoice] whereas [s-d] violates neither. This yields the wrong quantitative bias.

(d) [t-d] (O/E = 0.60) ≤ [s-d] (O/E = 0.55). (Discussed above.)

In addition, there are several phonotactic sequences that the grammar is unable to tell apart because no constraint distinguishes between them. This suggests that more constraints are needed.

(42) Pairs not distinguished by the grammar

(a) [t-n], [s-r], [s-l] (O/E = 0.70, 1.08, 1.13) are predicted to be equivalent.

(b) [n-l], [t-s], [r-n] (O/E = 0.32, 0.37, 0.56) are predicted to be equivalent.

(c) [t-l], [t-r], [s-n] (O/E = 0.78, 0.88, 1.17) are predicted to be equivalent.

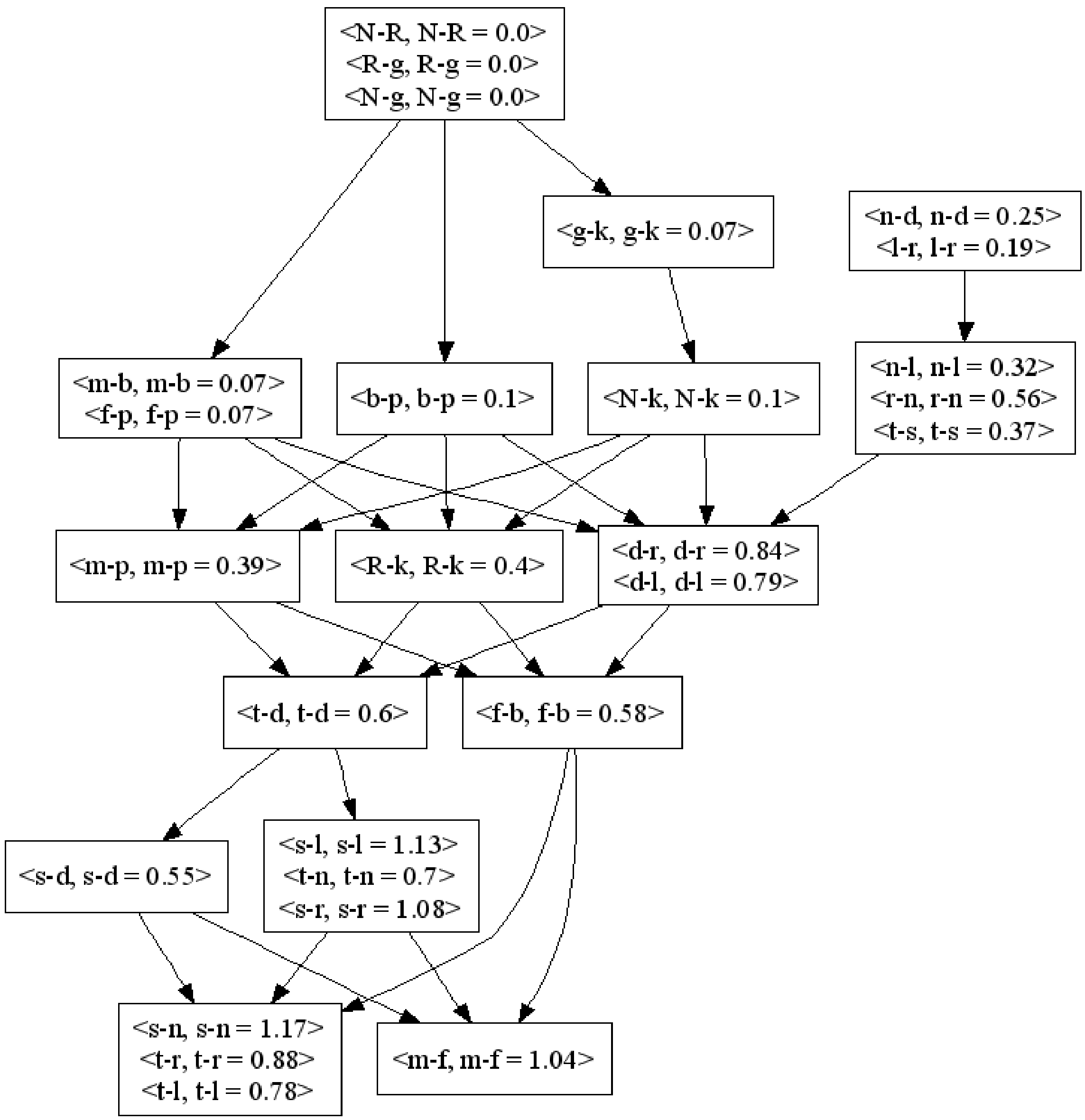
(d) [d-l], [d-r] (O/E = 0.79, 0.84) are predicted to be equivalent.

Finally, C&P 2006 rank the new constraints as shown in (43). They do not explicitly indicate where the new constraints belong in the existing stratum ordering. For this reason we simply use “Stratum X” and “Stratum Y”. This of course is not a flaw, but simply means that the ranking remains partial. The complete grammar with all the constraints and rankings yields the T-order in (44).

(43) The four-stratum constraint ranking for constraints on liquids

	Targeted Pairs	O/E values
Stratum X:		
OCP-cor[αcont, βvoice]	n-d, l-r	0.25, 0.19
Stratum Y:		
OCP-cor[αson, βvoice]	n-l, r-n, t-s	0.32, 0.56, 0.37
Stratum 4:		
OCP-cor[αson]	s-d, t-d	0.55, 0.60
Stratum 5:		
OCP-cor	t-l, t-n, d-l, d-r, t-r, s-l, s-r, s-n	0.78, 0.70, 0.79, 0.84 0.88, 1.13, 1.08, 1.17

(44) T-order with all rankings: precision = 0.901, recall = 0.735.



We now compare the success of the complete grammar for obstruents, nasals and liquids to the success of the earlier grammar for obstruents and nasals only. The inclusion of liquids turns out to lower both precision and recall:

(45) Precision and recall before and after liquids, with all rankings

(a)	Obstruents, nasals:	Precision = 0.986	Recall = 0.863
(b)	Obstruents, nasals, liquids:	Precision = 0.901	Recall = 0.735

We have already identified the problems that arise from the constraints. How about the problems that arise from the rankings? We can identify them by taking the T-order with rankings ((44)), the T-order without rankings ((39)), and by computing their difference. The result contains the arrows that originate from the rankings. There are 384 such arrows of which 11 go in the wrong direction. Eight of them arise because OCP-COR[αvoice] is ranked too high. Let us consider each of these cases in turn.

First, the sequences [d-r] and [d-l] violate the high-ranking OCP-COR[αvoice]. This incorrectly predicts that they should have lower O/E values than the sequence [f-b] ((46)) and the sequences [t-d], [t-n] and [s-d] ((47)). The actual O/E values are shown in (48).

(46) The problem: OCP-COR[αvoice] >> OCP-LAB[αson]

		OCP-COR [αvce]	FAITH	OCP-LAB [αson]	OCP-LAB	OCP-COR
d-r	d-r	*!				*
	☞ OTHER		*			
f-b	☞ f-b			*	*	
	OTHER		*!			

(47) The problem: OCP-COR[αvoice] >> OCP-COR[αcont], OCP-COR[αson]

		OCP-COR [αvce]	FAITH	OCP-COR [αcont]	OCP-COR [αson]	OCP-COR
d-r	d-r	*!				*
	☞ OTHER		*			
t-d	☞ t-d			*	*	*
	OTHER		*!			

(48) Predictions and O/E values (8 arrows):

(a)	d-r, d-l (O/E = 0.84, 0.79)	≤	f-b (O/E = 0.58)
(b)	d-r, d-l (O/E = 0.84, 0.79)	≤	t-n, t-d, s-d (O/E = 0.7, 0.6, 0.55)

Second, the grammar predicts that [n-d] and [l-r] should have identical O/E values. Tableau (49) shows why: the high-ranking OCP-COR[αcont, βvoice] suppresses the effect of the lower-ranking constraints that correctly distinguish between the two sequences. The actual O/E values are shown in (50).

(49) The problem: OCP-COR[αcont, βvce] >> OCP-COR[αson, βvce], OCP-COR[αson]

		FAITH	OCP-COR [αcont, βvce]	OCP-COR [αson, βvce]	OCP-COR [αvce]	OCP-COR [αcont]	OCP-COR [αson]	OCP-COR
n-d	☞ n-d		*		*	*		*
	OTHER	*!						
l-r	☞ l-r		*	*	*	*	*	*
	OTHER	*!						

(50) Predictions and O/E values (1 arrow):
n-d (O/E = 0.25) = l-r (O/E = 0.19)

Finally, the sequences [s-l] and [s-r] violate the high-ranking OCP-COR[αcont]. This incorrectly predicts that they should have lower O/E values than [m-f] which only violates the lower-ranking OCP-LAB. The actual O/E values are shown in (52). This is the only ranking problem not related to [voice].

(51) The problem: OCP-COR[αcont] >> OCP-LAB

		OCP-COR [αcont]	FAITH	OCP-LAB	OCP-COR
s-l	s-l	*!			*
	☞ OTHER		*		
m-f	☞ m-f			*	
	OTHER		*!		

(52) Predictions and O/E values (2 arrows):
s-l, s-r (O/E = 1.13, 1.08) ≤ m-f (O/E = 1.04)

We conclude that the constraints and rankings posited by C&P 2006 are clearly on the right track. However, some residual problems remain. The challenge is to improve C&P 2006's grammar so as to achieve better precision and recall. We will briefly explore some options in the following section.

4.3 *Revisions*

The above error analysis focused on discovering the sources of precision errors. We identified the high ranking of OCP-constraints related to [voice] as a source of several problems. How could we improve the analysis? The obvious move would seem to be to keep the [voice] constraints in the grammar, but leave them unranked. This will indeed help in terms of precision, but hurt in terms of recall. If we leave out all liquid-specific rankings, making OCP-COR[αcont, βvoice] and OCP-COR[αson, βvoice] unranked, the result is a slight gain in precision (precision = 0.903), but a slight loss in recall (recall = 0.726). If we further leave out OCP-COR[αvoice] unranked, the result is a further improvement in precision (precision = 0.922), but at the cost of a large drop in recall

(recall = 0.656). This illustrates the trade-off between precision and recall: it is easy to achieve high precision, but at the cost of poor recall, and vice versa. The challenge for future work on Muna phonotactics is to try to modify C&P 2006's constraints and rankings in ways that improve precision without simultaneously compromising recall.

The analysis discussed above did not assume any a priori rankings. C&P 2006, p. 25, suggest that OCP-LAB and OCP-DOR may be universally ranked above OCP-COR (Prince and Smolensky 1993/2004) and hint at an alternative formulation of general OCP-PLACE constraints. These assumptions are spelled out in (53) and illustrated in (54).

(53) Alternative assumptions

(a) OCP-DOR and OCP-LAB constraints are universally ranked higher than the corresponding OCP-COR constraints.

(b) The general constraints OCP-DOR, OCP-LAB, and OCP-COR are violated only if the specific constraints are not violated.

(54) Illustrating the alternative assumptions

	OCP-LAB [αson]	OCP-COR [αson]	OCP-LAB	OCP-COR
(a) f-b	*			
OTHER				
(b) s-d		*		
OTHER				
(c) m-f			*	
OTHER				
(d) s-n				*
OTHER				

We constructed a grammar that is identical to the earlier grammar for obstruents, nasals and liquids except that it conforms to the assumptions in (53). The T-orders turn out rather similar. The new analysis does slightly better on recall, but slightly worse on precision. The analyses are compared without rankings in (55) and with rankings in (56).

(55) The difference in precision and recall, no rankings

(a) Original grammar: Precision = 0.820 Recall = 0.229
(b) Modified grammar: Precision = 0.797 Recall = 0.131

(56) The difference in precision and recall, with rankings

(a) Original grammar: Precision = 0.901 Recall = 0.735
(b) Modified grammar: Precision = 0.892 Recall = 0.737

Finally, we note that Coetzee and Pater assume slightly different OCP-constraints for Arabic (P&C 2005) and Muna (C&P 2006). The two sets of constraints yield T-orders that are only partially similar. Both predict the ordering [t-d] ≤ [t-n] to be universal. This prediction indeed holds true in both Arabic and Muna. The constraints for Arabic predict three further universals: [t-s] ≤ [t-n] and [l-n] ≤ [t-n], which hold for both Arabic and Muna, and [t-d] ≤ [t-s], which holds for Arabic, but not for Muna. The constraints for Muna predict [l-n] = [t-s]. The O/E values are indeed close in Muna ([l-n], O/E = 0.32; [t-s], O/E = 0.37), but rather far off in Arabic ([l-n], O/E = 0.06; [t-s], O/E = 0.52). The constraint set that would achieve 100% precision simultaneously in both languages still remains to be found.

5. Conclusion

It is often observed that lexical items can be more or less well-formed depending on the phoneme combinations they contain. This phenomenon is called gradient phonotactics. We have put forward an explanation of gradient phonotactics based on Optimality Theory (Prince and Smolensky 1993/2004). Our proposal brings together two separate lines of inquiry within Optimality Theory: the work on gradient phonotactics initiated by Pater and Coetzee (P&C 2005, C&P 2006) and the work on ranking entailments initiated by Prince (2002a, 2002b, 2006a, 2006b).

Why do gradient phonotactic generalizations exist? Some phoneme combinations are more marked than others. How is this formalized in the grammar? Marked phoneme combinations are grammatically more complex than unmarked phoneme combinations. The goal of this paper has been to make these intuitions explicit. We have identified a structure that orders phonological mappings in terms of their grammatical complexity (T-orders) and proposed a particular interpretation of this structure (The Complexity Hypothesis):

(57) The Complexity Hypothesis: The probability of an <input, output> mapping is inversely correlated with its grammatical complexity.

We tested the Complexity Hypothesis using C&P 2006's analysis of Muna phonotactics. We showed that the hypothesis is by and large supported by the data and identified some residual empirical problems. At a more general level, we have seen that gradient phonotactics does not require new theoretical devices, such as gradient constraints, but follows from the elementary notions of Optimality Theory: constraints, rankings, and relations among rankings.

Acknowledgements

References

Albright, Adam. 2006. 'Gradient phonotactic effects: lexical? grammatical? both? neither?' Paper presented at the 80th Annual Meeting of the Linguistic Society of America, Albuquerque, NM.

Anttila, Arto. 2002. 'Morphologically Conditioned Phonological Alternations', *Natural Language and Linguistic Theory* 20, 1-42. Also on http://roa.rutgers.edu/.

Anttila, Arto. 2006. 'Variation and opacity', *Natural Language and Linguistic Theory* 24:893-944. Also on http://roa.rutgers.edu/.

Anttila, Arto. 2007a. 'Word stress in Finnish', Paper presented at the 81st Annual Meeting of the Linguistic Society of America, Anaheim, CA. Handout available at http://www.stanford.edu/~anttila/research/papers.

Anttila, Arto. 2007b. 'Prosodic constraints on constituent ordering', Paper to be presented at the 26th West Coast Conference on Formal Linguistics, University of California, Berkeley.

Anttila, Arto and Curtis Andrus. 2006. 'T-orders', Ms. (comes with software), available at http://www.stanford.edu/~anttila/research/torders/t-order-manual.pdf.

Anttila, Arto, Vivienne Fong, Stefan Benus, and Jennifer Nycz. in progress. 'Consonant clusters in Singapore English', Ms., Stanford University, Columbia University, and New York University.

Bailey, Todd M. and Ulrike Hahn. 2001. 'Determinants of wordlikeness: Phonotactics or lexical neighborhoods?' *Journal of Memory and Language* 44, 568-591.

van den Berg, René. 1989. *A Grammar of the Muna Language*, Foris, Dordrecht.

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

Berkley, Deborah Milam. 1994a. 'Variability and Obligatory Contour Principle Effects', in Beals et al., (eds.), *CLS 30: Proceedings of the 30th Annual Meeting of the Chicago Linguistic Society, Volume 2: The Parasession on Variation and Linguistic Theory*, Chicago Linguistic Society, Chicago. pp. 1-12.

Berkley, Deborah Milam. 1994b. 'The OCP and gradient data', *Studies in the Linguistic Sciences* 24:59-72.

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

Blumenfeld, Lev. 2005. 'Matching ictus and stress in Latin hexameter endings', Paper presented at the 80th Annual Meeting of the Linguistic Society of America, Albuquerque, NM.

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.

Coleman, John and Janet Pierrehumbert. 1997. 'Stochastic phonological grammars and acceptability', in *Computational Phonology. Third Meeting of the ACL Special Interest Group in Computational Phonology*, Association for Computational Linguistics, Somerset. pp. 49-56.

Frisch, Stefan A., Nathan R. Large, and David B. Pisoni. 2000. 'Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords', *Journal of Memory and Language* 42, 481-496.

Frisch, Stefan and Bushra Zawaydeh. 2001. 'The Psychological Reality of OCP-Place in Arabic', *Language* 77, 91-106.

Frisch, Stefan A., Janet B. Pierrehumbert, and Michael B. Broe. 2004. 'Similarity avoidance and the OCP', *Natural Language and Linguistic Theory* 22(1), 179-228.

Greenberg, Joseph. 1950. 'The patterning of root morphemes in Semitic', *Word* 6, 162-181.
Greenberg, J. H. and J. J. Jenkins. 1964. 'Studies in the psychological correlates of the sound system of American English', *Word* 20, 157-177.
Hammond, Michael. 2004. 'Gradience, phonotactics, and the lexicon in English phonology', *International Journal of English Studies*, 4(2), 1-24.
Hay, Jennifer, Janet Pierrehumbert, and Mary Beckman. 2004. 'Speech Perception, Well-Formedness, and the Statistics of the Lexicon', in J. Local, R. Ogden, and R. Temple, (eds.), *Papers in Laboratory Phonology VI*, Cambridge University Press, Cambridge UK, pp. 58-74. Also at http://www.ling.canterbury.ac.nz/people/hay.shtml.
Hayes, Bruce, Bruce Tesar, and Kie Zuraw. 2003. 'OTSoft 2.1, software package', available at http://www.linguistics.ucla.edu/people/hayes/otsoft/.
Inkelas, Sharon. 1998. 'The theoretical status of morphologically conditioned phonology: a case study of dominance effects', in Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 1997*, Kluwer, Dordrecht, pp. 121-155.
Inkelas, Sharon and Cheryl Zoll. 2003. 'Is grammar dependence real?', ROA-587.
Kruskal, J.B. 1983. 'An overview of sequence comparison: Time warps, string edits, and macromolecules', *SIAM Review* 25(2), 201-237.
Manning, Christopher D. and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*, MIT Press, Cambridge, Massachusetts.
McCarthy, John. 1988. 'Feature geometry and dependency: A review', *Phonetica* 45, 84-108.
McCarthy, John. 1994. 'The phonetics and phonology of Semitic pharyngeals', in Patricia Keating, (ed.), *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology* 3, Cambridge University Press, Cambridge, pp. 191-233.
Ohala, J. and M. Ohala. 1986. 'Testing hypotheses regarding the psychological manifestation of morpheme structure constraints', in *Experimental Psychology*, Academic Press, Orlando, FL, pp. 239-252.
van Oostendorp, Marc, and Frans Hinskens. 2007. 'Segmental inventories and T-orders: Evidence from Dutch dialects', Ms. Meertens Institute.
Orgun, Cemil Orhan. 1996. *Sign-Based Morphology and Phonology with special attention to Optimality Theory*, Ph.D. dissertation, Department of Linguistics, University of California, Berkeley, ROA-171.
Padgett, Jaye. 1995. *Stricture in Feature Geometry*. Dissertations in Linguistics, Center for the Study of Language and Information, Stanford, CA.
Pater, Joe and Andries Coetzee. 2005. 'Lexically Specific Constraints: Gradience, Learnability, and Perception', *Proceedings of the Phonology-Morphology Circle of Korea*, Seoul, Korea, pp. 85-119.
Pierrehumbert, Janet. 1993. 'Dissimilarity in the Arabic verbal roots', in A. Schafer, (ed.), *NELS 23: Proceedings of the North East Linguistic Society,* GLSA, University of Massachusetts, Amherst, pp. 367-381.
Prince, Alan. 2002a. 'Entailed Ranking Arguments', ROA-500.
Prince, Alan. 2002b. 'Arguing Optimality', ROA-562.
Prince, Alan. 2006a. 'Implication & Impossibility in Grammatical Systems: What it is & How to find it', ROA-880.

Prince, Alan. 2006b. ‘No More than Necessary: beyond the 'four rules', and a bug report’, ROA-882.

Prince, Alan and Paul Smolensky 1993/2004. *Optimality Theory: Constraint Interaction in Generative Grammar*, Blackwell Publishing, Malden, Massachusetts.

Raffelsiefen, Renate. 1999. ‘Phonological constraints on English word formation’, in Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 1998*, pp. 225-287.

Vitevitch, M. S., P. A. Luce, J. Charles-Luce, and D. Kemmerer. 1997. ‘Phonotactics and syllable stress: Implications for the processing of spoken nonsense words’, *Language and Speech* 40, 47-62.

Zamma, Hideki. 2005. ‘Predicting Varieties: Partial Orderings in English Stress Assignment’, ROA-712.