

To appear in *Journal of the Phonetic Society of Japan*.

**Gradient Well-Formedness in Harmonic Grammar:
Phonological Performance as a Window on Phonological Competence**

Andries W. Coetzee*

Abstract. In this paper, I argue that phonological performance data provide information about phonological competence. An adequate model of phonological competence must hence be able to account for significant patterns observed in phonological performance data. Based on two word-likeness experiments, I point out some typical properties of phonological performance data, and then show how these properties can be accounted for in Harmonic Grammar.

Keywords. Gradient well-formedness, Harmonic Grammar, Performance and Competence.

1. Introduction

Ever since the earliest years of generative grammar, the validity of linguistic performance as data on linguistic competence has been viewed with some suspicion. This was in no small part due to statements by Chomsky in some of his earlier foundational writings about the generative enterprise – statements such as the (in)famous quote below:

We must make a fundamental distinction between *competence* (the speaker-hearer's knowledge of his language) and *performance* (the actual use of language in concrete situations) ... Observed use of language ... may provide evidence as to the nature of [competence], but surely cannot constitute the actual subject matter of linguistics, if this is to be a serious discipline.

(Chomsky 1965:4)

It is noteworthy that Chomsky does not claim that performance data are without value for the study of competence. In fact, he makes quite the contrary claim, namely that performance data can provide evidence about the nature of competence. This more positive valuation of performance data has been the

foundation of the subfield of psycholinguistics, and data on sentence processing have been used for a long time to validate and develop generative syntactic theories. In recent decades, the use of psycholinguistic performance data has also become increasingly more important in generatively oriented phonology, and it is currently generally accepted as valid data about phonological competence by most phonologists.¹

In this paper, I accept that performance data collected via psycholinguistic experiments provide valid information about human phonological competence. With this assumption comes the added responsibility that our models of phonological competence must be able to account for significant patterns observed in these performance data. I will focus on a few of the typical properties of performance data collected through experiments on phonological processing, and consider what a model of phonological competence should look like in order to account for these properties. More specifically, I will focus on the following three properties of typical phonological performance data:²

- (i) *Gradient*. Grammatical forms and ungrammatical forms are usually differentiated in a way that approaches categoricalness. However, the data also show evidence of finer gradient distinctions within the sets of grammatical and ungrammatical forms – i.e. there are degrees of well-formedness (Coetzee to appear; Hayes 2000).
- (ii) *Global and cumulative*. Words are processed as a whole so that all parts of a word contribute to how it is processed. Multiple ungrammatical structures across a word add up to contribute to the degree of ungrammaticality of the word. Similarly, a word that contains a single ungrammatical structure but consists otherwise of well-formed structures, will receive a boost from its well-formed parts during processing (Hay *et al.* 2004; Ohala & Ohala 1986).
- (iii) *Frequency based*. How a word is processed depends partially on its frequency of use, as well as the frequency of use of its constituent parts (Bailey & Hahn 2001; Treiman *et al.* 2000).

If we accept psycholinguistic performance data as valid information about phonological competence, then any adequate model of phonological competence must be able to account for these three typical

properties of psycholinguistic performance data. I will argue that one implementation of Harmonic Grammar (henceforth HG; Pater 2009; Smolensky & Legendre 2006), namely *noisy* HG (Coetzee & Pater to appear) meets this requirement. The rest of this paper is structured as follows: In section 2, I review the results of two specific phonological processing experiments, and show how these experiments exemplify the three properties listed above. Section 3 forms the core of the paper, and presents an account of the data from section 2 in the model of noisy HG. Finally, in section 4, I evaluate the success of the noisy HG account, and compare it to other widely used models of phonological competence.

2. Restrictions on [sC_iVC_i]-words in English

2.1. Basic facts

English allows words of the form [stVt] (*state, stout, stoat*, etc.), but not of the form [skVk] or [spVp] (**spape, *skake*). This forms part of a larger restriction on identical consonants allowed in the onset and coda of a single syllable in English (**slale, *smame*, etc.; Browne 1981; Fudge 1969). The absence of [spVp]- and [skVk]-forms from the English lexicon is usually assumed to be due to the fact that they are ungrammatical in English (Baertsch & Davis 2003; Coetzee 2004, 2008; Davis 1982, 1989, 1991), an assumption to which I will ascribe here too.

The categorical absence of [spVp]- and [skVk]-forms also has a gradient reflex in the English lexicon. In general, syllables with [t] in the onset and coda (*toot, stunt*, etc.) are more frequent than syllables with [p] (*pope, pulp*) or [k] (*cake, click*) in the onset and coda. The data in Table 1 list the frequency of each of these three kinds of syllables in the CELEX syllable sub-corpus (Baayen *et al.* 1995). More generally, syllables with any (not necessarily identical) coronals in the onset and coda (*state, ten*) are more frequent than syllables with any labials (*pope, map*) or any dorsals (*cake, king*) in the onset and coda. However, all three types of syllables appear less often than expected by chance. Table 2 based on data reported by Berkley (2000), shows both the frequency difference between the three types of syllables and the general underrepresentation of all three types.

[Table 1 goes here.]

[Table 2 goes here.]

To the extent that the three observations about phonological processing data mentioned in §1 are correct, we should therefore expect to see the following with regard to the processing of [stVt], [skVk] and [spVp] forms:

- (i) *Gradient.* There should be a sharp distinction in processing of grammatical [stVt]-forms and ungrammatical [skVk]- and [spVp]-forms. However, we expect gradient differences in the processing of the two kinds of ungrammatical forms.
- (ii) *Global/cumulative.* Although [stVt]-forms are grammatical, they contain coronals in onset and coda position – a structure that is underrepresented in the English lexicon. We therefore expect that these forms will show some processing deficit due to this. Similarly, although [skVk]- and [spVp]-forms are ungrammatical, they contain subparts that are grammatical and observed in English. We therefore expect these forms to receive some processing boost due to this.
- (iii) *Frequency.* Due to the large frequency advantage of syllables with two [t]’s, we expect [stVt]-forms to have a large processing advantage over [skVk]- and [spVp]-forms. Syllables with two [k]’s also have a frequency advantage over syllables with two [p]’s, however this advantage is markedly smaller than the frequency advantage of syllables with two [t]’s generally. We therefore expect a moderate processing advantage of [skVk]-forms over [spVp]-forms.

2.2. Experimental results

In this section, I discuss the results of two experiments on the processing of [stVt]-, [skVk]-, and [spVp]-forms that confirm the expectations listed in §2.1 above. Due to space limitations, and since these experiments were reported in detail elsewhere (Coetsee 2004, 2008, to appear), I will not discuss the design or results of the experiments in detail. The reader is referred to the original published reports on these experiments for more detail.

Experiment 1: Word-likeness rating. In this experiment, native speakers of American English were presented with non-word tokens of the form [stVt], [skVk], and [spVp], and were asked to rate each token on a five point scale in terms of its word-likeness, with a score of [5] corresponding to a form that is a perfectly good word and a score of [1] to a form that is not at all word-like. The results of the experiment are presented graphically in Figure 1. Overall, [stVt]-forms were rated as more word-like than [skVk]-forms ($t(198) = 9.2, p < .001$) and than [spVp]-forms ($t(198) = 9.1, p < .001$). Although [skVk]-forms were rated better than [spVp]-forms on average, this difference did not reach significance ($t(198) = .9, p = .17$).

[Figure 1 goes here.]

Experiment 2: Comparative word-likeness. The same speakers participated in this experiment, and the same kinds of tokens were used. However, in this experiment, tokens were presented two at a time, and participants had the task of selecting the member of each pair that was most word-like. Token pairs included [stVt]~[skVk], [stVt]~[spVp], and [skVk]~[spVp]. Figure 2 shows the percent times that a token of each type was selected in each of these three types of token pairs. [stVt]-tokens were selected more often than [skVk]-tokens ($t(299) = 11.9, p < .001$), and [spVp]-tokens ($t(299) = 15.4, p < .001$). Although the difference was smaller in the [skVk]~[spVp]-pairs, it was still significant, with [skVk]-tokens being selected more often ($t(299) = 2.3, p = .01$).

[Figure 2 goes here.]

These results give evidence for all three of the properties of psycholinguistic phonological processing data mentioned in §1 above.

- (i) *Gradience.* First, both experiments give evidence for a large difference between grammatical [stVt]-forms and ungrammatical [skVk]- and [spVp]-forms. Within the set of ungrammatical forms, both experiments also show evidence of gradient differences with [skVk]-forms being

rated better than [spVp]-forms. This difference is smaller than the difference between grammatical and ungrammatical forms, as evidenced inter alia by the fact that the difference reached significance only in one of the two experiments.

- (ii) *Global/cumulative*. Although [stVt]-forms were overall rated best, they did not receive perfect ratings. In Experiment 1, they did not consistently receive ratings of [5], and in Experiment 2, they were not selected 100% of the time. This is evidence that, although they are grammatical, there are not perfect – they do contain some structures that are at least avoided (such as syllables with multiple occurrences of coronals – see Table 2). Similarly, although [skVk]- and [spVp]-tokens are ungrammatical, they did not receive the worst possible ratings. In Experiment 1, they did not consistently receive a score of [1], and when in competition with [stVt]-forms in Experiment 2, they were selected at least some of the time. This gives evidence for the fact that they get a boost in well-formedness from those parts of their structure that are grammatical and actually observed in English.
- (iii) *Frequency* (see Table 1 and Table 2). The large word-likeness advantage of [stVt]-forms over the other two types reflects the large frequency advantage of these forms in the English lexicon. Similarly, the small advantage of [skVk]-forms over [spVp]-forms reflects the smaller frequency advantage of the [skVk]-forms over [spVp]-forms.

These results are typical of most results from phonological processing experiments. Under the assumption that such data give information about the structure of phonological competence, any adequate model of phonological competence must be able to account for data of this type. In the next section, I show how these data can be accounted for in noisy HG.

3. Gradient well-formedness in Harmonic Grammar

Harmonic Grammar (HG; Pater 2009; Smolensky & Legendre 2006) is a close relative to Optimality Theory (OT). Like OT, HG is a constraint-based theory of grammar. However, unlike in OT where

constraints are ranked, constraints are weighted in HG. HG has a noisy implementation, similar to the noisy version of OT (Stochastic OT; Boersma & Hayes 2001). In noisy HG, the weight of each constraint is perturbed by a small, randomly distributed, positive or negative number each time that the grammar is used. Constraint violations are marked with negative integers. A well-formedness or harmony score, \mathcal{H} , is calculated for each candidate according to the formula in (1). The candidate with the highest \mathcal{H} -score is selected as the output.

$$(1) \quad \mathcal{H}(cand) = (w_1 + nz_1)C_1 + (w_2 + nz_2)C_2 + \dots + (w_n + nz_n)C_n$$

Where w_i is the weight of constraint i , nz_i the noise associated with that constraint, and C_i the number of violations of candidate *cand* in terms of constraint i , expressed in negative integers.

In order to develop an HG-grammar, it is therefore necessary to determine the weights associated with each constraint. A learning algorithm for noisy HG (Boersma & Pater 2008), related to the Gradual Learning Algorithm of Stochastic OT, has been implemented in *Praat* (Boersma & Weenink 2009). This algorithm takes as input a set of grammatical forms, represented according to the frequency with which each form appears in the corpus from which the input is taken. The algorithm then learns the weights for each constraint that would result in a grammar that best approximates the learning input.

3.1. A noisy HG-grammar of [sC_iVC_i]-forms in English

In order to develop a noisy HG-account of the [sC_iVC_i]-restriction in English, I assume the constraints in (2). See Coetzee (2008) for a motivation of these constraints.

- | | | |
|-----|--|---|
| (2) | *p/*k/*t | No [p]/[k]/[t] |
| | * _σ [s+stop] | No sequence of [s+stop] in syllable initial position |
| | * _σ [C _i ...C _i] | No syllable with an identical consonant in onset and coda |
| | MAX | No deletion |

To determine the weights associated with these constraints, I ran a learning simulation using the noisy HG learning algorithm implemented in *Praat*. The learning input file was based on the frequency with which each of these constraints are violated in the CELEX syllable sub-corpus (Baayen *et al.* 1995). Specifically, I found in this corpus all syllables that violate at least one of these constraints. All of these syllables were included in the learning input, in accordance with their frequency listed in the corpus. The tableau in (3) shows the frequency structure of the input file, as well as the constraints violated by each of the input candidates. A deletion candidate was included as a losing competitor for each learning input.

(3) Structure of the learning input file³

Syllable type	*t	*k	*p	* _σ [s+stop	* _σ [C _i ...C _i]	Max	Frequency
stXt	-2			-1	-1		2,590
skXk		-2		-1	-1		15
stX	-1			-1			16,983
skX		-1		-1			4,038
spX			-1	-1			7,013
tXt	-2				-1		3,606
kXk		-2			-1		800
pXp			-2		-1		264
tX	-1						247,978
kX		-1					106,772
pX			-1				72,841

The learning algorithm was run ten times on this learning input file, using *Praat*'s default settings. Due to the noisiness in the evaluation procedure, a slightly different grammar is learned each time. The average weights that were learned for each constraint across the ten learning simulations are taken as the weights for the constraints in the rest of this paper. These average weights are listed in (4), and the tableau

in (5) shows how this grammar evaluates each of the three kinds of non-words that were used as tokens in the experiments discussed in §2.2.⁴

(4) Average constraint weights across ten learning simulations

MAX	175.8
*p	86.4
*k	82.5
* _σ [C _i ...C _i]	63.8
*t	52.4
* _σ [s+stop	40.9

(5) Evaluation of the non-word tokens used in the experiments

	175.8	86.4	82.5	63.8	52.4	40.9		
	MAX	*p	*k	* _σ [C _i ...C _i]	*t	* _σ [s+stop	<i>H</i>	<i>A</i>
/stVt/ → stVt				-1	-2	-1	-209.5	18.7
sVt	-1				-1		-228.2	-18.7
/skVk/ → skVk			-2	-1		-1	-269.7	-11.4
→ sVk	-1		-1				-258.3	11.4
/spVp/ → spVp		-2		-1		-1	-277.5	-15.3
→ sVp	-1	-1					-262.2	15.3

A quick inspection of this tableau shows that the grammar has been learned successfully. An /stVt/-input is mapped faithfully onto [stVt], corresponding to the fact that [stVt] is a grammatical form in English, and that words with this form do actually exist in the language. Inputs of the form /skVk/ or /spVp/, on the other hand, are mapped onto an unfaithful deletion candidate, corresponding to the fact that

these forms are absent from English, and assumed to be ungrammatical. The reason for the difference in which /stVt/-inputs and /skVk/- and /spVp/-inputs is rated by the grammar comes from the weight differences assigned to the constraints that differentiate between these three types of inputs, namely *t, *k and *p. The learning input file contained 277,353 instances of [t], more than twice as many as of [k] (112,440), and more than three times as many as of [p] (80,382). These differences in frequency resulted in very different weights for the constraints violated by [t], [k] and [p] respectively, with *t having a significantly lower weight than either *p or *k. The consequence is that an [stVt]-candidate with two [t]'s have a much higher \mathcal{H} -score than an [skVk]-candidate with two [k]'s or an [spVp]-candidate with two [p]'s.

3.2. Gradient well-formedness of [sC_iVC_i]-forms

Since all candidates in HG tableaux have an associated numeric \mathcal{H} -score, comparing candidates within the tableau of a single output for their relative well-formedness is a straightforward matter – the \mathcal{H} -scores of different candidates can be compared directly. On the face of it, well-formedness comparisons across tableaux of candidates from different inputs can be done in the same manner, i.e. by merely comparing the \mathcal{H} -scores of candidates, as suggested by Keller (2006). However, as first pointed out by Boersma (2004), there is a problem with this simple assumption. \mathcal{H} -scores are relative to the tableau in which they are assigned. This is easiest to see when unfaithful candidates are considered – violations in terms of faithfulness constraints are assigned relative to some specific input so that a candidate that will violate a faithfulness constraint in one tableau will not necessarily also violate that faithfulness constraint in a different tableau.

This causes several problems with direct between-tableaux comparison of \mathcal{H} -scores, one of which can be illustrated with the tableau from (5) above: a non-output candidate from one tableau can have a higher \mathcal{H} -score than the output candidate from another tableau. In (5), the \mathcal{H} -score for the non-output candidate [sVt] (from input /stVt/; -228.2) is higher than the \mathcal{H} -score of the output candidates [sVk] (from input /skVk/; -258.3) and [sVp] (from input /spVp/; -262.2).⁵

A more serious consequence of such a direct between-tableaux comparison cannot be illustrated with the grammar learned above. It is also possible that a candidate that is not a possible word can receive a higher \mathcal{H} -score in one tableau than what a possible word receives in a different tableau. Under the direct comparison approach to gradient well-formedness, a possible word would then have a lower relative well-formedness than an impossible word. This problem is illustrated with the tableau in (6) below, based on a discussion in Coetzee and Pater (2008). This tableau represents the grammar of a language with final-devoicing, assuming a positional faithfulness analysis along the lines of Lombardi (1999). In this analysis, voiced consonants, which violate *VOICE, are only permitted when devoicing would result in a violation of IDENT[voice]_{Onset}. This faithfulness constraint only applies in onset position. In coda position, only general IDENT[voice] applies. Since this is a language with final devoicing, [pad] is not a possible word in this language. Yet, the candidate [pad] (from input /pad/) has a higher \mathcal{H} -score than the actually possible candidate [bam.bam] (from input /bambam/).

(6) Impossible [pad] has higher relative well-formedness than possible [bam.bam]

	3	2	1	\mathcal{H}	\mathcal{A}
	IDENT[voice] _{Onset}	*VOICE	IDENT[voice]		
/pad/ pad		-1		-2	-1
⚡ pat			-1	-1	1
/bambam/ ⚡ bam.bam		-2		-4	2
pam.bam	-1	-1	-1	-6	-2
pam.pam	-2		-2	-8	-4

Pater (2007) suggests as solution that an acceptability score, \mathcal{A} , be calculated for each candidate according to the formula in (7) (see also Coetzee & Pater 2008). This formula determines an \mathcal{A} -score for every candidate, relative to its own tableau, thereby overcoming the problems associated with direct between-tableau comparison of \mathcal{H} -scores.

$$(7) \quad \mathcal{A}(cand) = \mathcal{H}(cand) - \mathcal{H}(best-competitor)$$

Where *cand* is the candidate whose \mathcal{A} -score is being calculated and *best-competitor* is the most harmonic member from the same input as *cand*, and where $cand \neq best-competitor$.

The last column of the tableaux in (5) and (6) contain the \mathcal{A} -scores for all candidates in these tableaux, calculated according to the formula in (7). The first thing to note about these \mathcal{A} -scores is that they distinguish between the candidate selected as output and all other candidates from a tableau – the selected output always has a positive \mathcal{A} -score and all other candidates from the tableau have negative \mathcal{A} -scores. This follows directly from how \mathcal{A} -scores are defined in (7). The actual output will always have the highest \mathcal{H} -score in its tableau, so that its \mathcal{A} -score will be positive even after subtraction of the \mathcal{H} -score of a competing candidate. On the other hand, any non-output candidate will have a lower \mathcal{H} -score than its best competitor (the actual output), so that it is guaranteed to have a negative \mathcal{A} -score.

This property of \mathcal{A} -scores also directly solves the problems that were associated with direct comparison of \mathcal{H} -scores between tableaux. Consider first the problem that was pointed out with regard to the tableau in (5). Although the non-output candidate [sVt] (from /stVt/) has a higher \mathcal{H} -score than the actual outputs [sVk] (from /skVk/) and [sVp] (from /spVp/), [sVt] actually has a lower \mathcal{A} -score (-18.7) than both [sVk] (11.4) and [sVp] (15.3). The problem with regard to the tableau in (6) is similarly solved by the \mathcal{A} -scores. Although ungrammatical [pad] (from /pad/) has a higher \mathcal{H} -score than grammatical [bam.bam] (from /bambam/), [pad]'s \mathcal{A} -score (-1) is lower than that of [bam.bam] (2).

[Figure 3 goes here.]

Comparison of Figure 3 with Figure 1 and Figure 2 shows that the \mathcal{A} -scores correspond well to the way in which the participants in the experiments responded, and captures the three properties of phonological performance data discussed in §1.

- (i) *Gradience.* There is a large difference in \mathcal{A} -scores between possible words ([stVt]) and impossible words ([skVk] and [spVp]) – the difference between [stVt] and [skVk] is 30.1, and between [stVt] and [spVp] it is 34.0. This account therefore captures the large well-formedness difference between possible and impossible words. But, importantly, the \mathcal{A} -scores of the two types of impossible words are not equal. The \mathcal{A} -score of [skVk]-forms is 3.9 higher than that of [spVp]-forms. This corresponds with the subtle difference in ratings that participants in the experiments assigned to tokens of these types, and shows that the HG-account developed here can capture the subtle gradient well-formedness differences between different ungrammatical forms.
- (ii) *Global/cumulative.* The \mathcal{H} -scores, and the \mathcal{A} -scores that are calculated from \mathcal{H} -scores, take into consideration the performance of a candidate on all constraints. The whole candidate therefore contributes to its \mathcal{H} -score so that \mathcal{H} -scores are the result of global and cumulative evaluation. For instance, the specific \mathcal{A} -score of [stVt]-forms represented in Figure 3 is determined by the fact that these forms contains several structures that violate some constraints – it contains two [t]’s earning two violations of *t, it contains the sequence [tVt] earning it a violation of * $_{\sigma}[C_i \dots C_i]$, and it contains the sequence [st] earning it a violation of * $_{\sigma}[s+\text{stop}]$. The \mathcal{A} -scores therefore reflect global, cumulative evaluation of the forms.
- (iii) *Frequency.* The input file used during the learning simulation described in §3.1 reflects the frequency with which the different constraints are violated in English. The constraint weights learned therefore also encode this frequency information in the grammar. Since the \mathcal{H} -scores, and hence \mathcal{A} -scores, depend on the constraint weights, these scores also reflect the frequency information. This is clear from comparison of the structure of the learning input file from (3) with

Figure 3. The input file contained many more examples of [t] (277,353) than of [k] (112,440) or [p] (80,382). This is reflected in the weights of constraints *t (52.4), *k (82.5) and *p (86.4), and hence also in the \mathcal{A} -scores in Figure 3 – the \mathcal{A} -score of [t]-containing [stVt] is much higher than that of [k]-containing [skVk] and [p]-containing [spVp]. The smaller frequency difference between [k] and [p] in the input file is also reflected in the smaller weight difference between *k and *p, and in the smaller difference in \mathcal{A} -scores between [skVk] and [spVp].

4. Concluding discussion

In addition to noisy HG, other models of phonological grammar have also been used recently to account for gradient well-formedness. In this section, I briefly review two of these, arguing that noisy HG better accounts for the properties of phonological processing data pointed out above.

4.1. Stochastic OT

Boersma and Hayes (2001) propose an account of gradient well-formedness in Stochastic OT (see also Hayes 2000). In Stochastic OT, constraints are ranked on a continuous ranking scale, with every constraint having a specific basic ranking value on this scale. Every time that the grammar is used to evaluate output candidates for some input, the ranking value of each constraint is perturbed by a small normally distributed random amount of noise. The basic ranking value of a constraint and the noise associated with the constraint at a specific evaluation time are added together to determine the selection point of the constraint – the ranking position where the constraint will function during the particular evaluation occasion. Due to the addition of noise, the selection points of two constraints with basic ranking values that are close together may invert between consecutive evaluation occasions, potentially resulting in variation.

Assume that C_1 and C_2 are two constraints, with C_1 having a higher basic ranking value than C_2 . Since C_1 has a higher basic ranking value, the selection point of C_1 will also usually be higher than that of C_2 ($C_1 \gg C_2$). However, due to the contribution of noise during evaluation, the selection point of C_1 will

sometimes be lower than that of C_2 ($C_2 \gg C_1$). If these two possible rankings result in selection of different optimal candidates, variation will be observed. The probability of each of the two possible rankings depends on how close together the basic ranking values of C_1 and C_2 are – the closer they are, the more likely the inversion is to occur, is hence the more likely the ranking $C_2 \gg C_1$ will be observed. The frequency with which each of the possible variable outputs will be selected can therefore be controlled by controlling the distance between the basic ranking values of the constraints.

This model has been applied with great success to many examples of variable production. However Boersma and Hayes (2001) extend the model to account for gradient well-formedness. They propose that the well-formedness of some form is directly related to the likelihood of it being selected as output. If some input has two variable output forms [out₁] and [out₂], with [out₁] appearing more frequently than [out₂], then [out₁] will also be relatively more well-formed than [out₂].

Although this approach to relative well-formedness has been applied successfully to several examples in the literature, it has one serious shortcoming – it applies only to well-formedness differences between variant outputs. This follows directly from the way that well-formedness is calculated in the model – the well-formedness of some form is directly related to how likely it is to be selected as output. This model can therefore not account for any potential well-formedness difference between two invariant possible words. Since there is no variable phenomenon in English phonology that would apply to an input like /blik/, such an input would be mapped invariantly onto [blik]. Similarly, there is no variable process that would apply to an input like /blɪʒ/, and such an input would be mapped invariantly onto [blɪʒ]. The Stochastic OT account of gradient well-formedness would hence predict no well-formedness difference between such forms. This runs counter to typical results in the literature, that show that there are well-formedness differences between such forms, usually related to frequency – since [k] is a much more frequent segment than [ʒ] in English, [blik] is likely to be rated better than [blɪʒ] by English speakers (see the discussion on gradience and frequency in §1).⁶

For the same reason, this model of well-formedness can also not account for well-formedness differences between impossible (ungrammatical) forms. Impossible forms will never be selected as output.⁷ Returning to the examples discussed earlier in this paper, since both [skVk] and [spVp] are assumed to be impossible words of English, neither of them will ever be selected as output. They are therefore both expected to be absolutely, and equally, ill-formed. This, again, runs counter to results from the literature that shows clear evidence of well-formedness differences between different impossible forms, typically correlated with the frequency of the constituent parts of the impossible forms (see the discussion on gradience and frequency in §1).

4.2. Rank-ordered OT

Coetzee (2004, 2008, to appear) presents a different OT-based account of gradient well-formedness, called the “Rank-Ordered model of EVAL”, or ROE, for short (see also Berent *et al.* 2001; Everett & Berent 1998). ROE agrees with HG and differs from the Stochastic OT model in two respects: First, in ROE, an OT grammar does more than merely select the best (optimal) candidate from a candidate set, but rather imposes a well-formedness rank-ordering on the total candidate set. In HG, where every candidate is assigned an \mathcal{H} -score, all candidates can also be ordered in terms of their relative well-formedness. Secondly, in ROE, it is also possible to compare candidates between different tableaux (i.e. output candidates from different input forms). These two differences between ROE and Stochastic OT enable ROE to avoid the problems pointed out above for Stochastic OT. First, ROE rank-orders the whole candidate set, and can therefore account for well-formedness differences between all candidates in the candidate set, also for the ungrammatical, impossible forms. Secondly, since ROE allows for comparison between tableaux, it can also rank-order different outputs (possible words) for their well-formedness.

Coetzee (2008) proposes the OT grammar in (8) to account for the [sC_iVC_i]-restriction in English.⁸ The tableau in (9) shows how this grammar evaluates inputs of the kind /stVt/, /skVk/, and /spVp/. The grammar faithfully maps /stVt/ onto [stVt], showing that this form is rated as a possible word by the

grammar. On the other hand, /skVk/ and /spVp/ are unfaithfully mapped, showing that [skVk] and [spVp] are not rated as possible words.

(8) *spVp >> *skVk >> MAX >> *stVt

(9) /sC_iVC_i/-inputs in English according to Coetzee (2008)⁹

			*spVp	*skVk	MAX	*stVt
/stVt/	1	stVt				*
	2	sVt			*!	
/skVk/	2	skVk		*!		
	1	sVk			*	
/spVp/	2	spVp	*!			
	1	sVp			*	

The tableau in (10) shows the comparison between the candidates [stVt], [skVk], and [spVp] from the three different inputs. This tableau correctly captures the relative word-likeness ratings given by the participants in the experiments discussed above in §2.2. [stVt] was rated most word-like by the participants, and it also occupies the highest position on the rank-ordering imposed by the grammar in (10). Similarly, [spVp] was rated least word-like in the experiments, and it occupies the lowest position in the harmonic rank-ordering in (10).

(10) Gradient well-formedness of [stVt]-, [skVk]- and [spVp]-forms

		*spVp	*skVk	MAX	*stVt
1	/stVt/ → [stVt]				*
2	/skVk/ → [skVk]		*		
3	/spVp/ → [spVp]	*			

Although this account of relative well-formedness succeeds where Stochastic OT fails, it has other drawbacks when compared to the HG account. First, it has the same problems as discussed above with

regards to using \mathcal{H} -scores to compare the well-formedness of candidates across different tableaux (see §3.2). Specifically, a candidate that is an impossible word can be rated as relatively more well-formed than a possible candidate. This is shown in (11) using the same final devoicing grammar from (6) above. The impossible form [pad] is here also rated as more well-formed than the possible form [bam.bam].

(11) An impossible form is rated more well-formed than a possible form

		IDENT[voice] _{Onset}	*VOICE	IDENT[voice]
1	/pad/ → [pad]		*	
2	/bambam/ → [bam.bam]		**	

A second shortcoming of the ROE account of gradient well-formedness is that it imposes only a relative harmonic ordering on the candidates. HG, on the other hand, gives a numeric interpretation to the harmonic ordering so that the size of the well-formedness difference between two forms is relevant in HG. Concretely, HG can distinguish between the following two situations: (i) $\mathcal{A}(cand_1) = -100$, and $\mathcal{A}(cand_2) = -200$, (ii) $\mathcal{A}(cand_1) = -100$, and $\mathcal{A}(cand_2) = -101$. In both situations, HG would rate $cand_1$ as more well-formed than $cand_2$. However, since the \mathcal{A} -score of $cand_2$ is so much lower in (i) than in (ii), $cand_1$ will be rated as much more well-formed than $cand_2$ in (i) than in (ii). In ROE, these two situations cannot be distinguished. This shortcoming of ROE can be illustrated by comparing the ROE tableau in (10) with the HG tableau in (5). ROE correctly captures that [stVt] is more word-like than [skVk], and that [skVk] is more word-like than [spVp]. However, ROE does not capture the fact that the size in word-likeness difference in these two comparisons is vastly different (see the results of the experiments in §2.2). In the HG account in (5), on the other hand, this size difference is also captured. The \mathcal{A} -score of [stVt] is 30.1 higher than that of [skVk], while the \mathcal{A} -score of [skVk] is only 3.9 higher than that of [spVp].

4.3. Final evaluation

If phonological performance data are to be accepted as valid information about phonological competence, then our models of phonological competence must be able to account for significant generalizations observed in performance data. In this paper, I have focused on three specific typical properties of phonological performance data, and have shown that noisy HG can account for these properties. There is a large amount of data about phonological performance in the literature – both about phonological processing and about phonological production. This represents a source of information that has gone largely unnoticed by generatively oriented theoretical phonology over the past several decades. Exploring these data promises to lead to interesting new developments in our understanding of phonological competence, and will lead to interesting developments in phonological theory.

Endnotes

* Assistant Professor, Department of Linguistics, University of Michigan, Ann Arbor, MI 48109, USA.

¹ See Ohala (1986) for an early call to phonologists to take this kind of data seriously, and Coetzee et al. (2009) and Goldrick (to appear) for recent reviews, documenting the increased importance of this kind of data in phonology.

² These properties are documented in just about every paper that reports on some phonological processing experiment. The references cited here are therefore mere examples, and any number of alternative studies could have been cited in stead.

³ “X” stands for any segmental material so that [stXt] represents both words like ‘state’ [stert] with only a vowel between the two [t]’s and words like ‘straight’ [streit] with a vowel and a consonant between the [t]’s. The [skXk] syllables all represent syllables where the two [k]’s are separated by more than just a vowel – i.e. syllables like ‘skunk’ [skɹŋk] or ‘squeak’ [skwik]. The learning input file used is available from the author upon request.

⁴ An HG-tableau shares many properties with the more well-known OT-tableaux. Constraints are listed across the top, here in the order of descending weight rather than descending ranking as in an OT-tableau. The weight of each constraint is listed above the name of the constraint. Candidate output forms are listed along the left side of the tableau. Constraint violations are indicated by negative integers rather than by asterisks. The second to last column shows the \mathcal{H} -score for every candidate, calculated according to the formula in (1). In all tableaux in this paper, the noise associated with each constraint has been left off for the sake of simplicity. The candidate with the highest \mathcal{H} -score is selected as output, and is indicated by the familiar pointy hand that marks the output in an OT-tableau. Since \mathcal{H} -scores are negative, the highest \mathcal{H} -score is the one closest to zero. The final column gives the \mathcal{A} -score which is discussed in §3.2.

⁵ Note that [sVt] is a possible word, and would in fact be the output for an /sVt/-input. Both [sVt] and [sVk] are possible words according to the grammar developed above. The problem is hence not that an impossible word receives a higher \mathcal{H} -score than a possible word. It is more subtle: a form that is ungrammatical in one specific tableau ([sVt] from /stVt/) receives a higher \mathcal{H} -score than a form that is grammatical in another tableau ([sVk] from /skVk/). The more serious problem is also encountered: an impossible word (ungrammatical in all tableaux) can receive a higher \mathcal{H} -score than a possible word (grammatical in at least one tableau). This is shown in the next paragraph.

⁶ Hammond (2004) does extend Stochastic OT to cases of this kind. However, he does not provide an explanation for the absence of variation.

⁷ At least not in practice. In principle, all forms are possible outputs in Stochastic OT. Some forms just have an vanishingly small probability of being selected.

⁸ The *sCVC-constraints are locally conjoined versions of the markedness constraints used in the HG account developed above – see (2). See Coetzee (2008) for a motivation for the ranking between these constraints. See Pater (2007, 2009) for an explanation of why OT needs locally conjoined constraints while HG does not.

⁹ The harmonic ordering of candidates are indicated by Arabic numerals next to the candidates. The best candidate, marked by the numeral “1” is the optimal candidate of classic OT, and is hence selected as output.

References

- Baayen, Harold M., Richard Piepenbrock & Leon Gulikers. (1995). *The CELEX Lexical Database (CD-ROM)*. Philadelphia: Linguistic Data Consortium.
- Baertsch, Karen & Stuart Davis. (2003). The split margin approach to syllable structure. *ZAS Papers in Linguistics* 32, 1-14.
- Bailey, Todd M. & Ulrike Hahn. (2001). Determinants of wordlikeness: phonotactics or lexical neighborhoods? *Journal of Memory and Language* 44, 568-591.
- Berent, Iris, Daniel Everett & Joseph Shimron. (2001). Do phonological representations specify variables? Evidence from the Obligatory Contour Principle. *Cognitive Psychology* 42.
- Berkley, Deborah M. (2000). *Gradient Obligatory Contour Effects*. Northwestern University.
- Boersma, Paul. (2004). *A Stochastic OT Account of Paralinguistic Tasks Such as Grammaticality and Prototypicality Judgments*. Ms., University of Amsterdam. [ROA #648.].
- Boersma, Paul & Bruce Hayes. (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32, 45-86.
- Boersma, Paul & Joe Pater. (2008). *Convergence Properties of a Gradual Learning Algorithm for Harmonic Grammar*. Ms., University of Amsterdam, and University of Massachusetts, Amherst. [ROA #970.].
- Boersma, Paul & David Weenink. (2009). *Praat: Doing Phonetics by Computer (Version 5.1.20)*. [Computer Program.] Retrieved October 31, 2009, from <http://www.praat.org>.
- Browne, Wayles. (1981). Slavic -ba and English *slil: two persistent constraints. *Folia Slavica* 4, 219-225.
- Chomsky, Noam. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Coetzee, Andries W. (2004). *What It Means to be a Loser: Non-Optimal Candidates in Optimality Theory*. University of Massachusetts.
- Coetzee, Andries W. (2008). Grammaticality and ungrammaticality in phonology. *Language* 84, 218-257.

- Coetzee, Andries W. (to appear). Grammar is both categorical and gradient. In Parker, Stephen (ed.) *Phonological Argumentation*. London: Equinox.
- Coetzee, Andries W., René Kager & Joe Pater. (2009). Introduction: phonological models and experimental data. *Phonology* 26, 1-8.
- Coetzee, Andries W. & Joe Pater. (2008). Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language and Linguistic Theory* 26, 289-337.
- Coetzee, Andries W. & Joe Pater. (to appear). The place of variation in phonological theory. In Goldsmith, John, Jason Riggle & Alan Yu (eds.) *Handbook of Phonological Theory . 2nd Edition*. Cambridge: Blackwell.
- Davis, Stuart. (1982). On the internal organization of syllable constituents. In Larson, Thomas G. (ed.) *Coyote Papers 3: Working Papers in Linguistics of the University of Arizona*. Tucson: University of Arizona, 3-22.
- Davis, Stuart. (1989). Cross-vowel phonotactic constraints. *Computational Linguistics* 15, 109-110.
- Davis, Stuart. (1991). Coronals and the phonotactics of nonadjacent consonants in English. In Paradis, Carole & Jean-Francois Prunet (eds.) *The Special Status of Coronals: Internal and External Evidence*. San Diego: Academic Press, 49–60.
- Everett, Daniel & Iris Berent. (1998). *The Comparative Optimality of Hebrew Roots: An Experimental Approach to Violable Identity Constraints*. Ms. University of Pittsburg and Florida Atlantic University. [ROA #235.].
- Fudge, Erik C. (1969). Syllables. *Journal of Linguistics* 5, 253–286.
- Goldrick, Matt. (to appear). Utilizing psychological realism to advance phonological theory. In Goldsmith, John, Jason Riggle & Alan Yu (eds.) *Handbook of Phonological Theory . 2nd Edition*. Cambridge, MA: Blackwell.
- Hammond, Michael. (2004). Gradience, phonotactics, and the lexicon in English Phonology. *International Journal of English Studies* 4, 1-24.

- Hay, Jennifer, Janet Pierrehumbert & Mary E. Beckman. (2004). Speech perception, well-formedness, and the statistics of the lexicon. In Local, John, Richard Ogden & Rosalind Temple (eds.) *Phonetic Interpretation. Papers in Laboratory Phonology VI*. Cambridge: Cambridge University Press, 58-74.
- Hayes, Bruce. (2000). Gradient well-formedness in Optimality Theory. In Dekkers, Joost, Frank van der Leeuw & Jeroen van de Weijer (eds.) *Optimality Theory: Phonology, Syntax, and Acquisition*. Oxford: Oxford University Press, 88-120.
- Keller, Frank. (2006). Linear Optimality Theory as a model of gradience in grammar. In Fanselow, Gisbert, Caroline Féry, Ralph Vogel & Matthias Schlesewsky (eds.) *Gradience in Grammar: Generative Perspectives*. Oxford: Oxford University Press, 270-288.
- Lombardi, Linda. (1999). Positional faithfulness and voicing assimilation in Optimality Theory. *Natural Language and Linguistic Theory* 17, 267-302.
- Ohala, John J. (1986). A consumer's guide to evidence in phonology. *Phonology Yearbook* 3, 3-26.
- Ohala, John J. & Manjari Ohala. (1986). Testing hypotheses regarding the psychological manifestation of morpheme structure constraints. In Ohala, John J. & Jeri J. Jaeger (eds.) *Experimental Phonology*. Orlando: Academic Press, 239-252.
- Pater, Joe. (2007). *Cumulative Ill-Formedness in Typological and Experimental Data*. Handout from the workshop on *Experimental Approaches to Optimality Theory*, May 2004, University of Michigan, Ann Arbor. [Available at <http://people.umass.edu/pater/index-6.html>].
- Pater, Joe. (2009). Weighted constraints in generative linguistics. *Cognitive Science* 33, 999-1035.
- Smolensky, Paul & Géraldine Legendre. (eds.) (2006). *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar. Volume 1: Cognitive Architecture. Volume 2: Linguistic and Philosophical Implication*. Cambridge, MA: MIT Press.
- Treiman, Rebecca, Brett Kessler, Stephanie Knewasser, Ruth Tincoff & Margo Browman. (2000). English speakers' sensitivity to phonotactic patterns. In Broe, Michael B. & Janet Pierrehumbert (eds.) *Papers in Laboratory Phonology V: Acquisition and the Lexicon*. Cambridge: Cambridge University Press, 269-282.

Table 1. Frequency counts from the CELEX syllable sub-corpus

	Example	Frequency
...t...t...	<i>toot, stunt</i>	6195
...k...k...	<i>cake, skunk</i>	815
...p...p...	<i>pup, pump</i>	264

Table 2. Frequency counts based on Berkley (2000)

	Example	Observed	Expected	Observed/Expected
Coronal	<i>ten, toot</i>	1148	1271.5	.90
Dorsal	<i>cake, king</i>	81	113.0	.71
Labial	<i>pup, map</i>	118	207.9	.57

Figure 1. Results of word-likeness ratings experiment. Error bars show 95% confidence intervals.

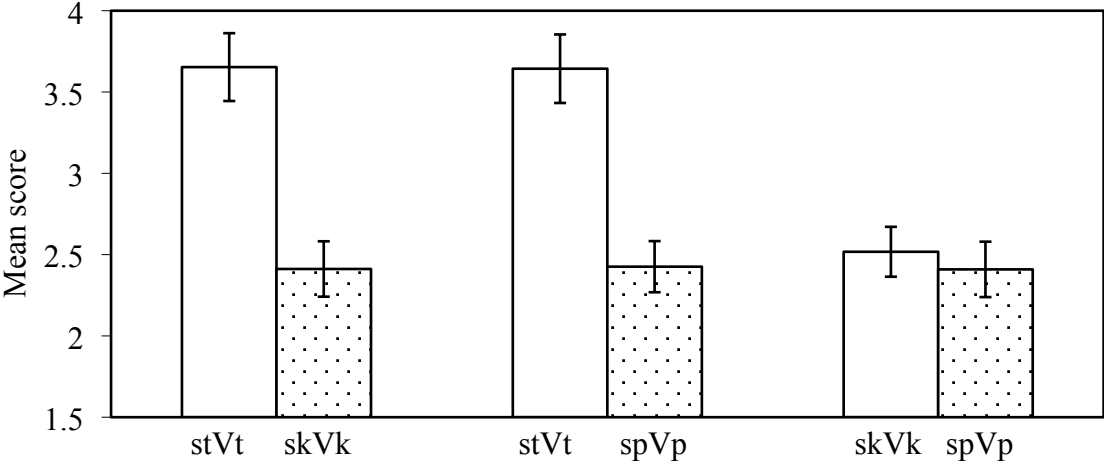


Figure 2. Results of comparative word-likeness experiment. Error bars show 95% confidence intervals.

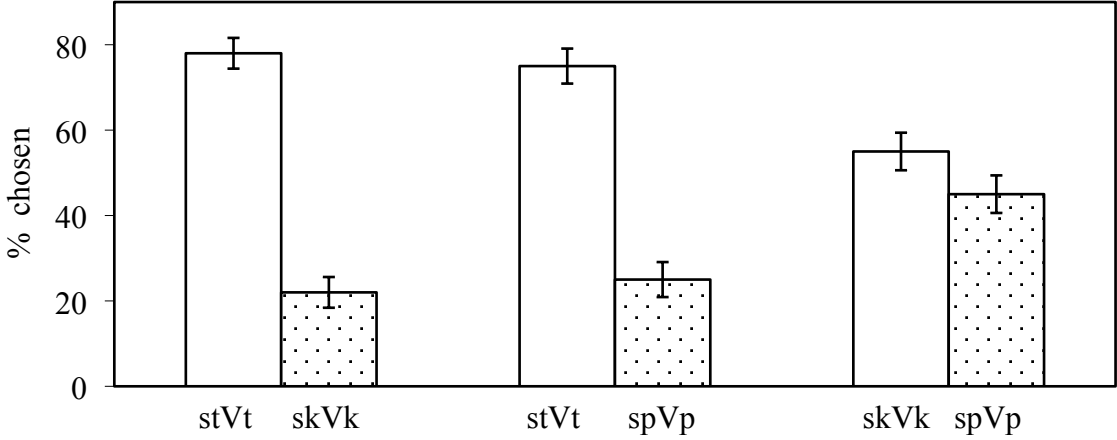


Figure 3. \mathcal{A} -scores for the non-word tokens used in Experiments 1 and 2

