Input Optimization: phonology and morphology*

Michael Hammond
U. of Arizona

Abstract

In this paper, I provide a unified account of three frequency effects in phonology. First, typologically marked elements are under-represented. Second, phonological changes are under-represented. Third, morphologically conditioned phonological changes are over-represented. These effects are demonstrated with corpus data from English and Welsh. I show how all three effects follow from a simple conception of phonological complexity. Further, I show how this notion of complexity makes predictions about other phenomena in these languages and that these predictions are borne out. I model this with traditional OT, but the proposal is consistent with any constraint-based formalism that weights constraints in some way.

1 Overview

In this paper I show that there are particular frequency effects governing the mapping from input to output. I demonstrate that, while they appear to conflict with each other, a simple unified account is possible. For this demonstration, a generic version of Optimality Theory (Prince & Smolensky, 1993; McCarthy & Prince, 1993) is assumed, but any constraint-based theory is compatible with the proposal.

I will demonstrate and provide a unified account for three statistical effects: i) the under-representation of marked phonological elements; ii) the under-representation of phonological changes; and iii) the over-representation of morphologically conditioned phonology.

The rarity of marked elements is well established. Typologically marked elements tend to be rarer than typologically unmarked elements in languages that have both. This applies to both marked elements and to marked configurations. The under-representation of phonological mappings between input and output is established by Hammond (2013): forms that undergo phonological changes between input and output are under-represented with respect to forms that do

*Thanks to Adam Albright, Elise Bell, Ricardo Bermudez-Otero, Amy Fountain, Chris Golston, S.J. Hannahs, Lionel Mathieu, Diane Ohala, and Maggie Tallerman for useful discussion. Thanks also to audiences at Manchester and Arizona and to the members of my Linguistics 514 class, Spring of 2015. Finally, thanks to three anonymous reviewers, an associate

not undergo changes. That there is over-representation of forms that undergo phonological changes conditioned by morphology is demonstrated by Hammond (2014). The latter paper provides the outlines of how this might be treated in the context of the under-representation effect. Here I put all these pieces together into an explicit account that also treats the typological effects and test it with a number of additional phenomena not previously treated.

The organization of this paper is as follows. I begin with classical frequency effects in the domain of typological markedness, reviewing data from English. The general phenomenon is that marked elements are less frequent than unmarked elements. Next, we turn to similar effects in the domain of phonological mapping, again using data from English. I show that phonological changes (*qua* faithfulness violations) are under-represented compared with non-changes. In the next section, I show that consonant mutation in Welsh exhibits the opposite skewing: changes induced by consonant mutation are *over-represented* compared with non-changes.

We next consider a variety of corpus data from English and Welsh, demonstrating that it is the morphological aspect of consonant mutation that causes this apparent different behavior and we provide an account of this difference. Finally, I conclude with a review of the general empirical results, theoretical claim, remaining questions, and directions for future research.

2 Typological markedness

In the following, I take typological markedness as an opposition between two elements a and b cross-linguistically. The element a is typologically marked with respect to b just in case a does not occur in a system unless b is there. In other words, the presence of a in a language implies the presence of b: $a \rightarrow b$ (Hammond et al., 1989).

It's well known that typologically marked elements tend to be less frequent than unmarked elements in the phonological systems that actually contain them.[1] For example, [d] is more marked typologically than [t] and, in systems that have both, [d] tends to be less frequent.[2] Marked phonological elements and configurations are avoided statistically in surface/output representations (Jakobson, 1968).

We can see this effect in English with word-initial coronal stops using the

editor, and the editors for additional feedback. All errors are my own.

[1]See, for example, Trubetzkoy (1939), Greenberg (1954), Greenberg (1966), Greenberg (1974), etc. See Maddieson (1984) for a compendious sample of such generalizations. See Berkley (1994a), Berkley (1994b), Frisch (1996), Frisch et al. (2000), Coetzee & Pater (2011), etc. for further discussion.

[2]If we look at the phonetic details, things can get much more complicated. For example, in a language like English, where /d/ is most often essentially voiceless, is it still more marked? The example in the text proceeds on either of two assumptions. One possibility is the traditional one: [t] and [d] are opposed in voicing at some level, and [d] is the more marked member of the pair. The other possibility is closer to the phonetics. The opposition is between [t^h] and [t], and [t] (orthographic *d*) is the more marked member of the pair (Vaux & Samuels, 2005).

Brown corpus (Kučera & Francis, 1967).[3] Voiced stops are more marked than voiceless stops typologically. This is evidenced by the number of languages that have voiceless stops, but not voiced stops, and the virtual absence of languages of the other sort: with voiced stops, but not voiceless stops. Focusing, for convenience, on word-initial position, what we find is that, in English, observed (O) voiced stops are rarer than voiceless stops. More specifically, if we assume they should be equally frequent, the occurring distribution is significantly different.

(1)

Word type	O	E	O/E
d...	30245	39994	0.756
t...	49743	39994	1.244

This distribution is significant: $X^2(1, N = 79988) = 4752.863$, $p < .001$.[4] One might doubt a comparison based on a written corpus, but we get the same effect with the spoken Buckeye corpus (Pitt et al., 2007). This corpus has 284732 words.

(2)

Word type	O	E	O/E
d...	9843	11194.5	0.879
t...	12546	11194.5	1.121

This distribution is also significant: $X^2(1, N = 22389) = 326.330$, $p < .001$.

There are similar effects with *phonotactic* or *contextual* markedness. For example, consonant clusters are more marked than singletons crosslinguistically; if a language has clusters, it will necessarily have singletons, but not vice versa. Correspondingly, if a language has clusters, they will be less frequent than the corresponding singletons. For example, in English once again, word-initial singleton [d] is more frequent that word-initial [dC] clusters in the Brown corpus and in the Buckeye corpus.

(3)

	Brown			Buckeye		
Type	O	E	O/E	O	E	O/E
dV...	27578	15039.5	1.834	9371	4875.5	1.922
dC...	2501	15039.5	0.166	380	4875.5	0.078

This distribution is also significant in the Brown corpus: $X^2(1, N = 30079) = 20906.810$, $p < .001$, and in the Buckeye corpus: $X^2(1, N = 9751) = 8290.235$, $p < .001$.

Prince & Smolensky (1993) show that a framework like OT can accommodate *systemic* markedness, implicational generalizations of the form: if a language has /d/, it will also have /t/. The explanation for this comes from the

[3]The Brown corpus is a fairly old written corpus of approximately one million words. I use it here because it is familiar to many and publicly available, allowing readers to more easily confirm the claims made here themselves.

[4]Statistical methods are described in detail in the appendix. This includes how χ^2 and expected values are calculated.

claims that: i) there is a universal set of constraints; and ii) that these constraints can interact only via strict ranking. On the assumption that we have a faithfulness constraint FAITH and a markedness constraint *d, it follows that only two kinds of phonological system are possible (4).

(4)

	Constraints	Phonological system
a.	FAITH ≫ *d	{t, d}
b.	*d ≫ FAITH	{t}
c.	impossible	{d}

One ranking gives us (4a), another gives us (4b), but there is no ranking of these two constraints that will produce (4c).

That said, orthodox OT provides no direct account of *statistical* markedness. We turn to this below.

3 Phonological changes

The distributional patterns discussed in the previous section extend to other parts of the phonology as well. Specifically, the same kinds of skewings apply at the phrasal level and apply to input-output mappings.

Marked phonological configurations can be repaired phonologically as well. These changes are also statistically avoided. An example is the Rhythm Rule (Liberman & Prince, 1977; Hammond, 1988; Hayes, 1984).[5] The Rhythm Rule refers to the phenomenon whereby a primary stress in English is shifted leftward onto a preceding secondary stress if it would otherwise occur too close to a following stress.

The chart separates these two factors, clash and whether there's a preceding secondary stress.

(5)

	2ndary	Clash	Example	
a.	yes	yes	ùnknówn mén	→ únknòwn mén
b.	yes	no	rétàil mén	
c.	no	yes	alóof mén	
d.	no	no	háppy mén	

In (5a) we see stress shifting leftward because the primaries are too close. In (5c) we see no shift, because there is no preceding secondary to shift the primary onto. In (5b) and (5d) we see no shift as the stresses are not close enough.

Hammond (2013) demonstrates that the 1st and 3rd cases above (5a,c) are statistically under-represented using the tagged Brown corpus plus CMU pronouncing dictionary.[6] The basic idea is to compare the distribution of these items in environments where the Rhythm Rule applies and those where it

[5] Years ago, Bolinger (1962) argued that clash is avoided in use in English. He doesn't show this statistically, but he was certainly the first to make the point.

[6] `http://www.speech.cs.cmu.edu/cgi-bin/cmudict`

doesn't. It's a little complex to do this because stress isn't marked in the tagged Brown corpus. It's also difficult because the environments where shift occurs depend on whether the relevant item is in a syntactic phrase with the following item and the stress of the first item is close enough to that of the second. Following Hayes (1984), stress shift aims for four-syllable intervals; hence two-syllable modifiers will be in the appropriate stress configuration if the following word has a stress on the first or second syllable. This is, of course, always true in English (Chomsky & Halle, 1968, etc.). The syntactic environment is approximated by comparing prenominal environments to all others. This isn't exact. For example, we might expect adjectives before other adjectives to constitute a Rhythm Rule environment and our search strategy groups these incorrectly. The idea is that the prenominal examples will be dominated by appropriate syntactic configurations for the Rhythm Rule, and examples of the second non-prenominal sort less so. This certainly isn't perfect, but it avoids having to do a full syntactic parse.

There are 1161192 words in Brown and 127008 words in CMU. There are 64028 adjective tokens in Brown and 8063 adjective types. Of these, 4049 occur in the CMU dictionary, and, of these, 1281 are disyllabic and can be analyzed.[7] The following chart just gives us the general pattern. As we might expect, there are a lot more trochaic adjectives than iambic. There are a lot more words with a single stress than two stresses.

(6)

Pattern	Example	Types	Tokens	Token freq.
$\acute{\sigma}\breve{\sigma}$	happy [hǽpi]	960	17921	0.87
$\breve{\sigma}\acute{\sigma}$	aloof [əlúf]	171	1920	0.09
$\acute{\sigma}\grave{\sigma}$	finite [fájnàjt]	85	422	0.02
$\grave{\sigma}\acute{\sigma}$	unknown [ʌnnón]	27	255	0.01

If we break these up into prenominal vs. non-prenominal tokens, we get the following.

(7)

Pattern	Example	Non-prenom.	Prenom.
$\acute{\sigma}\breve{\sigma}$	happy [hǽpi]	6785	11136
$\breve{\sigma}\acute{\sigma}$	aloof [əlúf]	970	950
$\acute{\sigma}\grave{\sigma}$	finite [fájnàjt]	118	304
$\grave{\sigma}\acute{\sigma}$	unknown [ʌnnón]	115	140

Overall, the distribution prenominally is significantly different from that non-prenominally, $X^2(3, N = 7988) = 270.205$, $p < .001$.

(8)

Pattern	Non-prenom. O	E	O/E
$\acute{\sigma}\breve{\sigma}$	6785	7099.311	0.956
$\breve{\sigma}\acute{\sigma}$	970	605.6345	1.6
$\acute{\sigma}\grave{\sigma}$	118	193.803	0.609
$\grave{\sigma}\acute{\sigma}$	115	89.2514	1.29

[7] One has to set aside 41 forms where the stress is incorrect: $\acute{\sigma}\acute{\sigma}$. Most are miscoded

This can be made more precise though. Two distributional patterns are important here. First, the distributions of items like *háppy* and *alóof* are significantly different with respect to prenominal and non-prenominal environments. In prenominal position, words like *aloof* represent 8% of adjectives with no secondary stress, while in non-prenominal position, they account for 13%. This difference is significant: $X^2(1, N = 7755) = 231.300$, $p < .001$. This shows us that unresolvable clash, a marked configuration, is under-represented.

(9)

Pattern	Non-prenom. O	E	O/E
σ́σ̆	6785	7145.431	0.95
σ̆σ́	970	609.5689	1.59

Second, the distributions of items like *fínìte* and *ùnknówn* are significantly different across prenominal and non-prenominal environments as well. In prenominal position, words like *unknown* represent 32% of adjectives with secondary stress, while in non-prenominal position, they account for 49%. This difference is significant: $X^2(1, N = 233) = 34.290$, $p < .001$. This shows that *resolvable* clash is also under-represented.

(10)

Pattern	Non-prenom. O	E	O/E
σ́σ̀	118	159.5315	0.74
σ̀σ́	115	73.46847	1.57

Let's confirm this pattern with the Buckeye corpus. First, I tagged the corpus with the Stanford part-of-speech tagger (Toutanova et al., 2003).[8] The procedure was then the same as above. We see the same general pattern in the basic distribution below.

(11)

Pattern	Example	Types	Tokens	Token freq.
σ́σ̆	happy [hǽpi]	420	4243	0.91
σ̆σ́	aloof [əlúf]	64	312	0.07
σ́σ̀	finite [fájnàjt]	28	87	0.02
σ̀σ́	unknown [ʌ̀nnón]	11	30	0.01

Prenominally vs. elsewhere in Buckeye, we get a similar distribution to what we saw in Brown:

(12)

Pattern	Example	Non-prenom.	Prenom.
σ́σ̆	happy [hǽpi]	2162	2081
σ̆σ́	aloof [əlúf]	221	91
σ́σ̀	finite [fájnàjt]	43	44
σ̀σ́	unknown [ʌ̀nnón]	20	10

Overall in Buckeye, the distribution prenominally is significantly different from that non-prenominally, $X^2(3, N = 2226) = 71.140$, $p < .001$, just as in Brown.

morphologically complex forms.

[8]Marked silences and disfluencies were treated as sentence breaks.

(13)

Pattern	Prenom. O	E	O/E
$\acute{\sigma}\breve{\sigma}$	2081	1967.544	1.06
$\breve{\sigma}\acute{\sigma}$	91	201.1226	0.452
$\acute{\sigma}\grave{\sigma}$	44	39.13246	1.12
$\grave{\sigma}\acute{\sigma}$	10	18.20114	0.549

As with the Brown data, two distributional patterns are important here. First, the distributions of items like *háppy* and *alóof* are significantly different with respect to prenominal and non-prenominal environments. In prenominal position, words like *aloof* represent 4% of adjectives with no secondary stress, while in non-prenominal position, they account for 9%. This difference is significant: $X^2(1, N = 2172) = 66.731$, $p < .001$. This shows us that unresolvable clash, a marked configuration, is under-represented in Buckeye as well.

(14)

Pattern	Prenom. O	E	O/E
$\acute{\sigma}\breve{\sigma}$	2081	1970.568	1.06
$\breve{\sigma}\acute{\sigma}$	91	201.4318	0.452

Second, as in Brown, the distributions of items like *fínìte* and *ùnknówn* are significantly different across prenominal and non-prenominal environments as well. In prenominal position, words like *unknown* represent 19% of adjectives with secondary stress, while in non-prenominal position, they account for 32%. This difference is significant: $X^2(1, N = 54) = 4.360$, $p = 0.037$. This shows that *resolvable* clash is also under-represented in Buckeye, as in Brown.

(15)

Pattern	Prenom. O	E	O/E
$\acute{\sigma}\grave{\sigma}$	44	36.85714	1.19
$\grave{\sigma}\acute{\sigma}$	10	17.14286	0.583

What we have then is that, in both the written Brown corpus and the spoken Buckeye corpus, unrepairable clash is under-represented, but so is repaired clash. What this means is that there's more going on than just that marked elements and configurations are avoided; phonological repair is also avoided.

Other explanations for these skewings are, of course, possible. One might suppose that the distribution of the four classes of adjectives is accidentally connected to the semantics. That trochaic adjectives tend to have meanings more appropriate for prenominal position, while iambic adjectives tend to meanings more appropriate for other positions. There are at least three reasons to reject this kind of approach as an explanation. First, showing that there is a statistical correlation between semantic or syntactic categories and phonological properties is not itself an explanation. What we need is some explanatory principle and/or some grammatical mechanism that makes the connection necessary, that makes it follow from general principles. Second, appeal to accidental semantic or syntactic biases is not a unified account. The account developed here involves a single explanatory principle that covers all cases. Finally, the account developed here is not only unified, but it's sensible. It extends existing grammatical machinery in a straightforward way. To characterize these phonological facts, I appeal to phonological machinery in a comprehensible way, rather than appealing to accidental semantic facts.[9]

4 Morphological processes: mutation

But the plot thickens. Let's now turn to a rather different phenomenon; here I show that Welsh mutation exhibits the *opposite* distribution from the English cases.

Let's review the general pattern. Welsh has three basic mutations. These are a class of consonantal changes that happen word-initially in a morphosyntactically prescribed set of environments. I'll focus on the soft mutation which makes the following changes.

(16)

p	→	b	b	→	v	ɬ	→	l
t	→	d	d	→	ð	r̥	→	r
k	→	g	g	→	∅	m	→	v

Other consonants do not change in this environment. I call the changing consonants *mutators*; [f, s, χ, n] etc. are *non-mutators*.

Let's look at a few examples to see how it works. In (17a), we see a feminine singular noun mutating after the definite article, and in (17b) we see that an adjective modifying a feminine singular noun will also mutate. The object of certain prepositions mutates (17c), as does the direct object of an inflected verb (17d).

(17)
a. fem. sing. noun after det.
   cath [**k**aːθ] 'cat'
   y gath [ə **g**aːθ] 'the cat'
b. adj. after fem. sing. noun
   dewr [**d**ɛwr] 'brave'
   cath ddewr [kaːθ **ð**ɛwr] 'brave cat'
c. element after prep. *i*
   Manceinion [**m**ankejnjɔn] 'Manchester'
   i Fanceinion [i **v**ankejnjɔn] 'to Manchester'
d. obj. of inflected verb
   tarw [**t**aru] 'bull'
   gwelodd hi darw [gwɛlɔð hi **d**aru] 'She saw a bull'

Hammond (2014) demonstrates the *opposite* effect for Welsh mutation from what we saw in the previous section. If we just look after prepositions that trigger the soft mutation vs. all other environments, we can see this difference. As mentioned above, certain prepositions in Welsh mark the following word with the soft mutation. These include the following:

[9] Thanks to an anonymous reviewer for extremely helpful discussion of these issues.

(18)

i	‘to’	[i]	o	‘from’	[ɔ]
dros	‘over’	[drɔs]	drwy	‘through’	[druj]
am	‘for’	[am]	ar	‘on’	[ar]
at	‘at’	[at]	dan	‘under’	[dan]
gan	‘by’	[gan]	heb	‘without’	[hɛb]
wrth	‘to’	[urθ]			

The CEG corpus (Ellis et al., 2001) is a publicly available tagged corpus of written Welsh containing 1223501 words. In addition, it gives the lemma form for all tokens. In this corpus, we see that in other environments, mutators constitute 0.212, but after prepositions that trigger soft mutation, mutators form 0.31 of the total.

(19)

	prep.	prep. freq.	non-prep.	non-prep. freq.
mutators	30405	0.31	239108	0.212
non-mutators	67779	0.69	886209	0.788

This difference is significant: $X^2(1, N = 98184) = 5542.824$, $p < .001$.

(20)

	prep. O	E	O/E
mutators	30405	20862	1.46
non-mutators	67779	77322	0.877

This means that while we avoid unresolvable stress clash in English, and while we avoid resolvable stress clash in English, the *opposite* is true for Welsh soft mutation.

This is surprising, so let’s dig deeper here and make sure of it. Personal names in Welsh do *not* undergo any of the mutations. This is not true for native and nativized geographic names, which can undergo the mutations, e.g. *i Fanceinion* [i **v**ankejnjɔn] ‘to Manchester’ in (17c) above.

(21)

Not mutated		Mutated	
i Pedr	'to Peter'	i ben	pen 'head'
[i pɛdɛr]		[i bɛn]	[pɛn]
gan Tomos	'by Thomas'	gan dad	tad 'father'
[gan tɔmɔs]		[gan dad]	[tad]
am Catrin	'about Catherine'	am gi	ci 'dog'
[am katrɪn]		[am gi]	[ki]
at Bethan	'toward Bethan'	at fws	bws 'bus'
[at bɛθan]		[at vʊs]	[bʊs]
heb Dafydd	'without David'	heb ddyn	dyn 'man'
[hɛb davɨð]		[hɛb ðɨːn]	[dɨːn]
wrth Gerallt	'to Gerald'	wrth ŵr	gŵr 'husband'
[urθ gɛraɬt]		[urθ uːr]	[guːr]
o Mair	'from Mary'	o foch	moch 'pigs'
[o majr]		[o vɔx]	[mɔx]
dan Llinos	'under Llinos'	dan lif	llif 'flood'
[dan ɬinɔs]		[dan li(v)]	[ɬi(v)]
ar Rhys	'on Rhys'	ar ran	rhan 'part'
[ar r̥ɨs]		[ar ran]	[r̥an]

Let's consider how often personal names begin with mutatable consonants. If mutation is avoided—like rhythm and clash in English—we would expect names to begin with mutatable consonants more often than non-names. In fact, the opposite is the case, consistent with the reversal we saw above in mutation contexts for non-names: names are less likely to begin with a mutatable consonant.[10]

(22)

	non-names	freq.	names	freq.
mutators	117961	0.486	6066	0.218
non-mutators	124630	0.514	21775	0.782

This distribution is significant: $X^2(1, N = 27841) = 8027.046$, $p < .001$.

(23)

	names O	E	O/E
mutators	6066	13538	0.448
non-mutators	21775	14303	1.52

[10]It's not that names necessarily avoid starting on mutable consonants, but that the distribution of mutable consonants is different between names and non-names with names showing fewer mutable initial consonants and non-names showing more. The facts presented are consistent with the other interpretation as well, that non-names prefer mutable consonants.

An intriguing possibility suggested by an anonymous reviewer is that, if both distributions are skewed, then what is being avoided is violations of a general rule that says: mutate in mutation contexts. More mutable consonants with non-names increases the application of this; less mutable consonants with names has a complementary effect, of avoiding the possibility of mutation in names. To test for this we need to know what the unskewed distribution is and whether both names and non-names depart from this. This is an empirical question that we leave for future research.

Yet another possibility suggested by a reviewer is to think of this in reverse. It's not that there is under-representation of mutators with names because they don't undergo mu-

One might be concerned that this is a written corpus and the patterns could be different in the spoken register. In fact, we observe a similar distribution in a spoken corpus as well. The Siarad corpus[11] is a transcribed spoken corpus of approximately 607450 words. This corpus is not tagged for part of speech, but the basic soft mutation comparison above can be approximated. I use only those prepositions triggering soft mutation that can be identified unambiguously, leaving aside *i* and *o*, which are ambiguous between prepositions and pronouns. I then search for all words that begin with sounds that unambiguously either could mutate or could be mutated, setting aside vowel-initial words since they can either be the mutated result of a [g]-initial word or a true vowel-initial word. This gives us the counts below, which can be compared to those in (19).

(24)

	prep.	prep. freq.	non-prep.	non-prep. freq.
mutators	1980	0.29	161703	0.269
non-mutators	4850	0.71	438917	0.731

Words beginning with mutatable consonants are more likely after mutating prepositions. This difference is smaller than in the written corpus, but is significant as well: $X^2(1, N = 6830) = 14.833$, $p < .001$.[12] Hence, we observe the same effect in the spoken register as well.

(25)

	prep. O	E	O/E
mutators	1980	1839	1.08
non-mutators	4850	4991	0.972

I conclude that mutation indeed exhibits the opposite distribution from the English cases we considered in the previous section.

5 A story

In this section I provide an analysis for this. Before proceeding, what has been established empirically?

First, under-representation of words like *ălóof* in prenominal position, [d] vs. [t] word-initially, [d] vs. [dr] word-initially, etc. shows that marked elements and configurations are statistically avoided.

Second, under-representation of words like *ùnknówn* in prenominal position shows that the Rhythm Rule, a phonological change, is statistically avoided as well.

On the other hand, Hammond (2014) shows that there is over-representation of mutatable consonants in contexts for mutation in Welsh, the opposite from

tation, but the other way around. They are an exception to mutation *because* of the under-representation. This seems quite possible, but doesn't change the fundamental observation above: the distribution of mutators in names and non-names is significantly different.

[11] Deuchar et al. (2014), available on-line at `http://www.siarad.org.uk`

[12] We cannot easily check the effect with personal names using the Siarad corpus as this corpus does not mark personal names, as in the CEG corpus.

what we saw in English. This was confirmed here by looking at non-names vs. personal names in Welsh in the CEG corpus, and showing the same reversal, and by looking at the same basic pattern in the spoken Siarad corpus, again showing the same basic pattern.

The first two cases above look a bit like Lexicon Optimization and it would be reasonable to try to build an account based on that machinery.[13] Let's then look back at what that is. Here's the original definition:

(26) **Lexicon Optimization** (Prince & Smolensky, 1993):
Suppose that several different inputs $I_1, I_2, \ldots, I_n$ when parsed by a grammar G lead to corresponding outputs $O_1, O_2, \ldots, O_n$, all of which are realized as the same phonetic form Φ—these inputs are all *phonetically equivalent* with respect to G. Now one of these outputs must be the most harmonic, by virtue of incurring the least significant violation marks: suppose this optimal one is labelled O_k. Then the learner should choose, as the underlying form for Φ, the input I_k.

The basic idea is that if there are multiple ways to produce an output form consistent with the facts of a language, choose the input that produces the fewest constraint violations.

To see this in action, let's consider a simple example. Imagine we have nasal place assimilation. We have a constraint against NC sequences with different place values which outranks the relevant faithfulness constraints. For heteromorphemic examples, we would have tableaux like this:

(27)

	/ɪn-pjur/	NC	IO-Faith
☞	ɪmpjur		!
	ɪnpjur	*!	

In the case above, we have an input /n/ which is realized as [m] before a labial. Because the example is heteromorphemic, we can assume that there are other contexts—perhaps vowel-initial—where we can determine that the prefix-final consonant is indeed /n/. However, there are tautomorphemic cases where the input is unknown. Imagine we have an output form [lʌmp]; this is consistent with several possible inputs—/lʌmp/ and /lʌnp/. Either input produces the same output.

(28)

	/lʌmp/	NC	IO-Faith
☞	lʌmp		
	lʌnp	*!	

	/lʌnp/	NC	IO-Faith
☞	lʌmp		*
	lʌnp	*!	

In these cases, Lexicon Optimization militates for the input that produces the desired output most harmonically. We can see this in a *reverse tableau*,

[13]Note that this is not an endorsement of Lexicon Optimization; we're simply using it as

where inputs are given along the left and the violations marked are those for the optimal candidate given that input.[14] Basically, insofar as possible, lexicon optimization insures that what you see is what you get.

(29)

	[lʌmp]	NC	IO-Faith
☞	/lʌmp/		
	/lʌnp/		*

There are, of course, no empirical consequences to Lexicon Optimization *by itself*. In fact, it is defined to apply only when there are no consequences. Let's see if it's profitable to view the under-representations we see in English as statistical analogs to Lexicon Optimization.

To accommodate the effects we saw in English, we need to expand the notion of lexicon optimization to accommodate comparisons between inputs *when the outputs are not the same*. To do this, let's first define a notion of *Phonological complexity* that applies to individual input–output pairings, but also to entire phonological systems. (The basic logic of this is that the complexity of a phonological system is proportional to the number of asterisks in its tableaux.)

First, define the output/surface forms of a language as a possibly infinite set.

(30) $O = \{O_1, O_2, \ldots, O_n\}$

Every member of that set has a corresponding (optimal) input form:

(31) $I = \{I_1, I_2, \ldots, I_n\}$

There is also, of course, for any phonology, a finite sequence or vector of constraints:

(32) $C = \langle C_1, C_2, \ldots, C_n \rangle$

Any input–output pairing, (I_i, O_i), then defines a finite vector of violation counts, some number of violations for each constraint earned by the winning candidate for that input.

(33) $\langle n_{C_1}, n_{C_2}, \ldots, n_{C_n} \rangle$

With these notions, Phonological Complexity (PC) is defined as follows:

---

inspiration. See Nevins & Vaux (2007) for discussion of some possible shortcomings of Lexicon Optimization.

[14]We include the markedness constraint NC here for completeness; markedness constraints are not determinative in reverse tableaux.

(34) **Phonological Complexity** (PC)
The phonological complexity of some set of forms is defined as the vector sum of the constraint violation vectors for surface forms paired with their respective most optimal inputs.

To produce a *relative* measure of PC given some set of n surface forms, divide the PC score for those forms by n.

This can be exemplified with our hypothetical nasal assimilation example again. Let us assume the following set of forms we wish to compute the PC for. Given the inputs provided in column 2, we have the constraint violations for winning candidates in columns 3 and 4.

(35)

	Input	Output	NC	IO-Faith
a.	/on pi/	om pi		*
b.	/an ba/	am ba		*
c.	/un bo/	um bo		*
d.	/en do/	en do		
e.	/on ta/	on ta		
f.	/un ti/	un ti		
g.	/an ku/	aŋ ku		*
h.	/in ga/	iŋ ga		*
i.	/on ke/	oŋ ke		*
			0	6

The relative complexity of this first system is: $\langle 0, 6\rangle/9 = \langle 0, 0.67\rangle$. We can compare the system in (35) with the one in (36). Here we have a different array of output forms, but the same logic for inputs and constraint violations.

(36)

	Input	Output	NC	IO-Faith
a.	/on pi/	om pi		*
b.	/an ba/	am ba		*
c.	/en do/	en do		
d.	/on ta/	on ta		
e.	/un ti/	un ti		
f.	/in di/	in di		
g.	/an ku/	aŋ ku		*
h.	/in ga/	iŋ ga		*
			0	4

The relative complexity of this second system is: $\langle 0, 4\rangle/8 = \langle 0, 0.5\rangle$.

The second system is less complex than the first:

(37) $\langle 0, 0.5\rangle < \langle 0, 0.67\rangle$

It would be reasonable to assume that more complex complexity vectors should be compared using the logic of strict ranking. Thus, for example:

(38) $\langle 0.9, 0.5\rangle > \langle 0.4, 0.67\rangle$

In the example above, the relative magnitudes for the higher-ranked constraint determines the relative complexity of the systems, rather than the relative magnitudes of the lower-ranked constraint.

The proposal then is that all phonological systems are skewed to be less complex.

(39) **Input Optimization**
All else being equal, phonological inputs are selected that minimize the phonological complexity of the system.

This alters the frequency of input–output pairings; it does not change the input representation of any particular form.

This works out just fine for the English cases. First, in the case of word-initial [t] vs. [d], we minimize the latter to avoid voicing markedness violations. Second, in the case of word-initial [d] vs. [dr], we minimize the latter to avoid markedness violations (*COMPLEX). Third, for prenominal *háppy* vs. *alóof*, we minimize the latter to avoid markedness violations (*CLASH). Finally, for prenominal *fínìte* vs. *ùnknówn*: minimize the latter to avoid faithfulness violations.

Let's go through each of these to show this. First, consider word-initial [t] vs. [d]. We assume there is a constraint penalizing voiced stops which we'll simply call *VCD-STOP. Imagine we have a sample of 100 words that begin with coronal stops with the following distribution:

(40)

Type	Count	FAITH	*VCD-STOP
t...	50	0	0
d...	50	0	50

The total PC score here is $\langle 0, 50\rangle$, and the relative PC score is $\langle 0, 50\rangle/100 = \langle 0, .5\rangle$. We can imagine a skewed distribution, of the sort we saw in English, but more extreme, like the following:

(41)

Type	Count	FAITH	*VCD-STOP
t...	75	0	0
d...	25	0	25

The total PC score here is $\langle 0, 25\rangle$, and the relative PC score is $\langle 0, 25\rangle/100 = \langle 0, .25\rangle$. The latter distribution, with fewer word-initial instances of [d], is thus less complex. The actual occurring and expected distributions from the Brown corpus, along with relative PC scores are given below:

(42)

	Observed	Expected
t–d	49743–30245	39994–39994
Rel. PC	$\langle 0, 0.38\rangle$	$\langle 0, 0.5\rangle$

The exact same logic applies in the case of word-initial [d] vs. [dr], except the relevant markedness constraint is *COMPLEX. A distribution like (43) is dispreferred to one like (44).

(43)

Type	Count	FAITH	*COMPLEX
d...	50	0	0
dr...	50	0	50

(44)

Type	Count	FAITH	*COMPLEX
d...	75	0	0
dr...	25	0	25

As in the previous pair, the relative PC score for the less preferred distribution is $\langle 0, 50\rangle/100 = \langle 0, .5\rangle$, while that for the preferred distribution is $\langle 0, 25\rangle/100 = \langle 0, .25\rangle$. The latter distribution, with fewer word-initial instances of [dr], is less complex. The actual distribution and relative PC scores for the Brown corpus are given below:

(45)

	Observed	Expected
dV–dC	27578–2501	15039.5–15039.5
Rel. PC	$\langle 0, 0.08\rangle$	$\langle 0, 0.5\rangle$

Prenominal *háppy* vs. *alóof* works exactly the same way with respect to the markedness constraint *CLASH. Here, the higher-ranked constraint is not a faithfulness constraint, since we know stress shift is generally possible in English, but a constraint that requires that if stress shifts, it shifts to a syllable that would otherwise bear secondary stress. For convenience, we call this SECONDARY. The first of the following distributions is less preferred than the second.

(46)

Type	Count	2NDARY	*CLASH
háppy	50	0	0
alóof	50	0	50

(47)

Type	Count	2NDARY	*CLASH
háppy	75	0	0
alóof	25	0	25

The math is exactly the same. Actual values and relative scores from Brown are given below.

(48)

	Observed	Expected
háppy–alóof	11136–950	10574.28–1511.724
Rel. PC	$\langle 0, 0.08\rangle$	$\langle 0, 0.87\rangle$

Finally, consider the case of prenominal *fínìte* vs. *ùnknówn*. Here what is ruled out is application of the Rhythm Rule, not clash per se. We can assume that when stress shift applies, it violates some version of OO-correspondence, a constraint requiring stress in a clash context to be the same as stress in other contexts. That constraint, in turn, is dominated by *CLASH, and of course by 2NDARY.

(49)

Type	Count	*CLASH	OO-Correspondence
fínìte X	50	0	0
únknòwn X	50	0	50

(50)

Type	Count	*CLASH	OO-Correspondence
fínìte X	75	0	0
únknòwn X	25	0	25

Again, the same math applies. Following are the true values and relative scores from Brown.

(51)

	Observed	Expected
fínìte–ùnknówn	304–140	224.8584–219.1416
Rel. PC	$\langle 0, 0.32\rangle$	$\langle 0, 0.51\rangle$

What about the Welsh examples? On the face of it, these look like Welsh is skewed so as to make its system *more* complex. Recall that in a mutation context, like after a preposition like *i* 'to', we get more instances of mutating consonants than in non-mutation contexts. Let's assume that there is a constraint that forces mutation in various environments; we can call it MUTATE. This constraint outranks the relevant faithfulness constraint. We get exactly the wrong prediction when we consider the same two hypothetical distributions as in the previous cases. Compare mutating items like *cath* [kaːθ] 'cat' vs. non-mutating items like *afal* [aval] 'apple' after *i*. First, a neutral distribution:

(52)

Type	Count	MUTATE	FAITH
i afal	50	0	0
i gath	50	0	50

What we would expect is fewer instances of constructions like *i gath*:

(53)

Type	Count	MUTATE	FAITH
i afal	75	0	0
i gath	25	0	25

We would then have $\langle 0, .25\rangle$, rather than $\langle 0, .5\rangle$.

The problem is that we get just the opposite. In mutation contexts, we get more instances of constructions like *i gath*. Schematically:

(54)

Type	Count	MUTATE	FAITH
i afal	25	0	0
i gath	75	0	75

We have $\langle 0, .75\rangle$, rather than $\langle 0, .5\rangle$, exactly the opposite of what is predicted by Input Optimization (39). Actual values and relative scores from the CEG corpus are given below.

(55)

	Observed	Expected
mutators–non-mutators	30405–67779	20862.19–77321.81
Rel. PC	$\langle 0, 0.69\rangle$	$\langle 0, 0.21\rangle$

Why might Welsh mutation behave in this way? The apparent difference is that mutation is a morphologically conditioned phonological change, so it seems reasonable to build an explanation on that difference. We can accommodate this under the Input Optimization rubric if, in fact, there is a constraint militating *for* the expression of morphological categories. The logic is that the reason why mutatable consonants are over-represented where they are is because there is a constraint that militates for morphological categories to be expressed.

The key point is that mutation, whether it be phonological, morphological, or lexical, must be subject to a constraint forcing morphological categories to be expressed. If mutation is, as described above, a morphologically conditioned phonological change, there's no issue. Some researchers have argued that mutation systems should be treated morphologically or lexically, either in terms of some special class of morphological rules or in terms of listed allomorphs. If one of these is correct, then application of that morphological rule or selection of allomorphs must be subject to a constraint that requires morphology to be expressed. I'll continue to describe mutation as a phonological process, but the general Input Optimization story developed here is consistent with other views of mutation as well.

In fact, Kurisu (2001) proposes something close to what we need:

(56) Realize Morpheme (RM)
Let α be a morphological form, β be a morphosyntactic category, and $F(\alpha)$ be the phonological form from which $F(\alpha{+}\beta)$ is derived to express a morphosyntactic category β. Then RM is satisfied with respect to β iff $F(\alpha + \beta) \neq F(\alpha)$ phonologically.

Soft mutation expresses morphological information. To the extent that a word in a soft mutation context begins with a mutatable consonant, violations of RM are avoided. Thus when a form like *cath* [kaːθ] undergoes soft mutation to become *gath* [gaːθ], RM is satisfied. When in soft mutation context, *afal* [aval] does not change, RM is violated.

If we add RM to the constraint set for Welsh and rank it above Faith, this accommodates both Welsh cases. Consider first mutable vs. nonmutable consonants in mutation contexts, the schematic example just considered. First, we have the case where mutators and non-mutators are relatively evenly distributed:

(57)

Type	Count	Mutate	RM	Faith
i afal	50	0	50	0
i gath	50	0	0	50

Note that MUTATE is here for completeness. RM forces the category to be expressed. Higher-ranked MUTATE forces the precise expression of that category.

Now the case where we have proportionally more mutators:

(58)

Type	Count	MUTATE	RM	FAITH
i afal	25	0	25	0
i gath	75	0	0	75

When relative PC is calculated with RM in the mix, we find the latter distribution is preferred: $\langle 0, .5, .5\rangle > \langle 0, .25, .75\rangle$. This is, of course, also true with the actual distribution in the CEG corpus where the occurring distribution $\langle 0, 0.31, 0.69\rangle$ is preferred to the expected distribution $\langle 0, 0.79, 0.21\rangle$. Notice that ranking, strict or otherwise, is key here. If RM is not ranked higher than FAITH, we do not get the desired effect.

The effects of Input Optimization are thus contingent on the ranking or weighting of the constraints in the language. Though the claim is that all languages will exhibit skewing to satisfy Input Optimization, it does *not* follow that all languages will skew in the same way. Different weights or rankings will entail different skewings. Consider for example, the common loss of final syllables, even when they may be desinential, marking inflectional properties of the word in question. This is a purely phonological process that is not conditioned by the morphology. How is such a thing possible on the story we are telling here? Presumably there is a high-ranked/weighted constraint that militates for the loss of such syllables and presumably it outranks RM. Input Optimization will minimize violations of the higher-ranked/weighted constraints over those of lower-ranked constraints like RM. See Section 8 below for more discussion.

Consider now nonmutable consonants in personal names vs. non-names: non-names begin with mutators more often than names do. If we take the distribution of mutators in names as the neutral distribution and the distribution with non-names as the distribution to be explained, this emerges directly: non-names have more mutators because that avoids violations of RM, just as in the examples just considered.

The RM constraint, however, is too restrictive. It would seem to imply that expression of a morphological category is minimal, that if it is already expressed elsewhere, there is no pressure. This, in turn, predicts that if mutation were to be triggered by an overt affix, then we should not see an over-representation effect.[15] In fact, such cases do occur in Welsh and they should show an over-representation effect as well.

There is a set of prefixes that trigger the soft mutation in Welsh, e.g. *cyn-* [kɨn/kən] 'ex-', *gor-* [gɔr] 'over-', *ail-* [ajl] 're-', *di-* [di] '-less', *hunan-* [hɨnan] 'self-', *is-* [is] 'sub-', *gwrth-* [gurθ] 'anti-', *cyd-* [kɨd/kəd] 'co-', *ad-* [ad] 're-', etc. Following are some of these, along with examples:

[15] This problem has been noted before (Ussishkin, 2000; Wolf, 2007).

(59)	cyn-	cyn-gleifion	‘ex-patients’	cleifion	‘patients’
		[kənglejvjɔn]		[klejvjɔn]	
		cyn-fyfyrwyr	‘ex-students’	myfyrwyr	‘students’
		[kənvəvərujr]		[məvərujr]	
		cyn-athro	‘ex-teacher’	athro	‘teacher’
		[kənaθro]		[aθro]	
	gor-	gor-lenwi	‘overfill’	llenwi	‘fill’
		[gɔrlɛnwi]		[ɬɛnwi]	
		gor-hoff	‘overfond’	hoff	‘fond’
		[gɔrhɔf]		[hɔf]	
		gor-glyfar	‘overclever’	clyfar	‘clever’
		[gɔrgləvar]		[kləvar]	
	ail-	ail-osod	‘replace’	gosod	‘place’
		[ajlɔsɔd]		[gɔsɔd]	
		ail-iaith	‘L2’	iaith	‘language’
		[ajljajθ]		[jajθ]	
		ail-fyw	‘relive’	byw	‘live’
		[ajlvɨw]		[bɨw]	

The examples above include stems that begin with mutators and stems that begin with non-mutators. What is the distribution? Is it like what we see after prepositions or like what we see elsewhere? To test this, I found all instances of these prefixes in the CEG corpus when marked with a hyphen and then did counts on the following stems.

One small complication is that the hyphen is not generally required for these prefixes. I chose to count the ones marked with overt hyphens as it's of course easier to find these in the corpus. However, the hyphen is required just in case we might have an orthographic ambiguity. This occurs when the final letter of the prefix and the first letter of the stem could be misparsed as part of the digraphs *ll* [ɬ], and *dd* [ð]. Thus, for example, a form like *ail-lenwi* [ajllɛnwi] ‘refill’ must be spelled with a hyphen to avoid the double letters being misinterpreted as *[ajɬɛnwi]. Including these items would bias our counts in favor of mutators, so items of this sort were excluded. (This slightly biases the count against mutators.) We find the following distribution which we compare with the distribution of mutation in the non-preposition environment (19) from CEG. I take the latter to be the default.

(60)

	prefix	prefix freq.	non-prep.	non-prep. freq.
mutators	833	0.7628205	239108	0.212
non-mutators	259	0.2371795	886209	0.788

There is a significant difference: $X^2(1, N = 1092) = 1976.534$, $p < .001$.

(61)

	prefix O	E	O/E
mutators	833	232	3.59
non-mutators	259	860	0.301

The effect is so large that we might worry something else is going on, that word-internal mutation is subject to other pressures not yet considered, but similar effects have been found in Welsh for plural suffixation and various associated stem vowel changes (Anderson, 2015). At this point, we must conclude that the pressure to express some morphological category via some phonological process is not contingent on whether that category might also be expressed elsewhere by an independent word, like a preposition, or by another morpheme. In the case at hand, the relevant morphological category is expressed by both a prefix, e.g. *ail-*, and the soft mutation. What's key is that the soft mutation doesn't apply to the prefix itself, but to the following stem. As it stands, RM would not enforce both operations since the prefix and the mutation are both in the same word. The RM constraint must therefore be revised so as to allow this. The key is to restrict the notion of "morphological form" in (56) to just a morpheme:

(62) Realize Morpheme (*revised*) (RM')
Let α be a *morpheme*, β be a morphosyntactic category, and $F(\alpha)$ be the phonological form from which $F(\alpha + \beta)$ is derived to express a morphosyntactic category β. Then RM' is satisfied with respect to β iff $F(\alpha + \beta) \neq F(\alpha)$ phonologically.

The revision is minimal and handles all the cases we've treated so far, including the prefix example just considered. In the prefix case, there are two domains for RM': the prefix itself and the stem. For a form like *ail-fyw* [ajlvɨw] above, we have *ail* [ajl] $\neq \emptyset$ and *fyw* [vɨw] $\neq$ *byw* [bɨw].

6 Confirmation

The solution developed in the previous section describes the cases we have considered in straightforward fashion, but we are relying on the assumption that it is *morphology* that behaves differently. It could just be that *Welsh* and *English* behave differently. In this section, this other possibility is ruled out by considering cases of morphologically triggered phonology in English and non-morphologically triggered phonology in Welsh.

Let's first look at an example in Welsh that's not connected to mutation. This example involves devoicing of voiced stops in the final coda of Welsh adjectives when they occur medially in comparatives and superlatives. The following examples show the basic form of comparatives and superlatives.

(63)

Stem		Comparative	Superlative
cyflym	[kəvlɨm]	cyflymach	cyflymaf
	‘fast’	[kəvləmax]	[kəvləma(v)]
llawn	[ɬawn]	llawnach	llawnaf
	‘full’	[ɬawnax]	[ɬawna(v)]
tawel	[tawɛl]	tawelach	tawelaf
	‘quiet’	[tawɛlax]	[tawɛla(v)]
twp	[tup]	twpach	twpaf
	‘stupid’	[tupax]	[tupa(v)]
trist	[triːst]	tristach	tristaf
	‘sad’	[tristax]	[trista(v)]

If the stem ends in a voiced stop, it devoices in this context:

(64)

Stem		Comparative	Superlative
gwlyb	[gʷliːb]	gwlypach	gwlypaf
	‘wet’	[gʷləpax]	[gʷləpa(v)]
caled	[kalɛd]	caletach	caletaf
	‘hard’	[kalɛtax]	[kalɛta(v)]
parod	[parɔd]	parotach	parotaf
	‘ready’	[parɔtax]	[parɔta(v)]
enwog	[ɛnwɔg]	enwocach	enwocaf
	‘famous’	[enwɔkax]	[ɛnwɔka(v)]
pwysig	[pujsɪg]	pwysicach	pwysicaf
	‘important’	[pujsɪkax]	[pujsɪka(v)]

This is an unusual process, the reverse of the more usual sort of voicing alternation one might see in an opposition like this: final devoicing. The historical analysis of these is that, at some point, the suffixes could be analyzed as *-hax and *-hav and the devoicing we see here is the residue of the effects of the [h]. Regardless of the history, the synchronic analysis must include some constraint or set of constraints that force this devoicing and our interest is in whether Faith violations are minimized here by Input Optimization.

This is a non-morphological process in the sense that it does not mark a morphological category. Specifically, the comparative and superlative are marked by the affixes above and devoicing here is simply restricted to certain morphological contexts. See Section 8 below for more discussion.

Let’s now consider the distributions.[16] It turns out that word-final voiceless stops are really rare, so more accurate comparisons can be made if we use a different category as our comparison base: nasals. The CEG corpus is a written one and there is an ambiguity in the Welsh orthography in terms of how to interpret *ng* [ŋ,ŋg], so we set it aside and only look at non-dorsals, therefore comparing the distribution of stem-final [b,d] vs. [m,n]. What we see is that voiced stops are *under-represented* in comparatives and superlatives.

[16]This cannot be tested with the Siarad corpus since, as already noted, that corpus is not

(65)

	Adj.	freq.	Comp./Sup.	freq.
Voiced stops	8940	0.473	72	0.351
Nasals	9946	0.527	133	0.649

This difference is significant: $X^2(1, N = 205) = 12.269$, $p < .001$. This establishes that Welsh and English are not *generally* reversed. Hence Welsh adjectives behave like other English phonological examples.

(66)

	comp./sup. O	E	O/E
Voiced stops	72	97	0.742
Nasals	133	108	1.23

We can look in the other direction as well. What about *morphological* cases in English? If the Input Optimization with RM' approach is correct, we expect them to behave like the Welsh soft mutation examples. English doesn't have anything like mutation, so we look at *morphological haplology* (Stemberger, 1981; Zwicky, 1987). One example is the genitive plural in English: the key fact is that overt plurals do not cooccur with the genitive.

(67)

man	man's	cat	cat's	
[mæn]	[mænz]	[kæt]	[kæts]	
men	men's	cats	cats'	*cats'(e)s
[mɛn]	[mɛnz]	[kæts]	[kæts]	*[kætsəz]

Another example is the suffix *-ly* which marks adverbs: the key fact is that it is not added to an adjective that already ends in *ly*.

(68)

Adjective	Adverb	
routine	routinely	
happy	happily	
weekly	weekly	(*weeklyly)

What we find in the Brown corpus is precisely what we would predict under Input Optimization with RM': forms like *cats'* in the genitive plural are statistically under-represented.

(69)

	men	cats
Non-genitive	4126	50984
Genitive	74	183

This difference is significant: $X^2(1, N = 4200) = 232.399$, $p < .001$.

(70)

	men O	E	O/E
Non-genitive	4126	4185	0.986
Genitive	74	15	4.93

tagged for part of speech.

Similarly, adverbs are much more frequent with adjectives that don't already end in *-ly* in Brown.

(71)

	routinely, etc.	weekly, etc.
Adjective	63080	948
Adverb	13922	3

This difference is also significant: $X^2(1, N = 951) = 202.629$, $p < .001$.

(72)

	weekly O	E	O/E
Adjective	948	779	1.22
Adverb	3	172	0.0174

One final example can be added here: word-final t/d-deletion. This is a well-known phenomenon initially studied by Guy (1991) and more recently by Turton (2012) and Coetzee & Kawahara (2013). The basic effect is that a word-final [t,d] can be deleted word-finally in English, e.g. in *friend* [frɛnd, frɛn]. The process is governed by a number of factors including whether the [t,d] appears in a cluster, whether the following word begins with a vowel, speech rate, informality, lexical frequency, etc. The relevant factor for us is that the process applies less readily if it would delete a consonant that is the sole exponent of the *-ed* past tense. Thus, all else being equal, we expect deletion to apply more readily to a word like *text* [tɛkst, tɛks] than a word like *boxed* [bakst, baks].

This is indeed the case in the Buckeye corpus. The following chart shows the relative retention of final [t,d] as a function of whether the word in question ends in *-ed*.

(73)

	-ed	Elsewhere
Retained	3857	41161
Deleted	799	19332

This distribution is significant: $X^2(1, N = 4656) = 468.807$, $p < .001$.

(74)

	-ed O	E	O/E
Retained	3857	3168	1.22
Deleted	799	1488	0.537

The facts of [t,d] deletion are quite consistent with the story here and support the hypothesis that a skewing reversal occurs when RM' would apply. We would expect deletion to be under-represented just in case it would violate RM' and that is what we see here. The Input Optimization story here would then be an alternative to the rule-based and constraint-based stratal approaches of Guy (1991) and Turton (2012) respectively.

Hence adjective devoicing in Welsh, the genitive plural in English, adverbs in English, and deletion of [t,d] in English work just as one would predict if the relevant distinction is morphological expression vs. phonological generalizations.

Since adjective devoicing in Welsh is not a morphological operation like lenition, it does not incur violations of RM'. Therefore we minimize faithfulness violations and expect under-representation of forms that would otherwise undergo devoicing. The genitive plural in English is an overt affix, thus clearly a morphological operation that RM' governs. Hence, we expect under-representation of the haplological cases as we find. Adverbs in English work the same way. RM' militates for the adverbial suffix to be expressed, so we expect to find under-representation of the haplological cases. In the case of deletion of [t,d], we see a case where a normal phonological process is limited by RM'.

7 How does Input Optimization work?

We've established a number of frequency effects that can all be unified and accommodated under the principle of Input Optimization (39), but how does Input Optimization work concretely? Here we address two questions. First, where does Input Optimization take place? Is it a part of grammar or something else? Second, wherever it may "live", why doesn't it overpower the rest of the grammar? The ideas in this section are extremely speculative, but are intended to lay the groundwork for future research.

We need to clarify two important aspects of the proposal to do these. First, Input Optimization does not entail that all languages work the same way. We've seen that it works to minimize constraint violations across the language and that it's sensitive to constraint ranking or weighting. Given that violations of higher-ranked or weighted constraints will be minimized over violations of lower-ranked or weighted constraints, and given that weights/ranking is at least partially language specific, it follows that the effects of Input Optimization will differ across languages.

Second, Input Optimization is a global effect, beyond the lexicon. We've seen a number of cases where Input Optimization might be taken as an effect in the lexicon, some mechanism by which the number of words that fit some phonological requirement are more or less than expected. However, two facts militate against an exclusively lexical story. First, all of our counts have been corpus counts, not dictionary counts. That is, we're explicitly considering how often words and constructions are used, rather than how often words occur in a dictionary. Second, as just noted, we've also seen a number of cases where it's phrases or multi-word patterns that are skewed. Assuming phrases are not generally listed lexically, this argues against attributing Input Optimization exclusively to the lexicon. One might counter that the statistical combinatory properties of lexical items can be stored in the lexicon, and this is certainly true, but this amounts to extending our notion of the lexicon to include statistical syntactic properties.

Given that Input Optimization extends beyond the lexicon, there are at least four ways we might think of it: i) as an historical effect; ii) as a property of acquisition; iii) as a performance constraint; or iv) as evidence for a different kind of phonological architecture. The first two are related, as are the last two.

I treat each of these four in turn.

Input Optimization could be a property of historical change specifically. That is, there is pressure for historical changes to selectively reduce the phonological complexity (34) of the system as a whole. The basic idea is that Input Optimization is a mechanism of historical change and the effects we've seen are not enforced by the grammar, but the result of historical accretion. This is a reasonable approach. Historical change is often a byproduct of the acquisition process, so we would have to carefully distinguish this from a purely acquisition-based story. (See below.) We would also need to think carefully about the phrasal skewings we've seen. We would need to allow for historical changes that change how often various words might cooccur.

Another possibility, somewhat related to the historical approach, is to view Input Optimization as a property of acquisition. The basic idea here is that the acquisition process is biased to minimize phonological complexity. Again, the effects we see would be a consequence of changes that occur during acquisition, not enforced by the adult grammar per se. If this were true, this would certainly have consequences in the historical domain, but we could in principle distinguish the two views. There are historical changes that occur with adults. If Input Optimization is an acquisition effect, then we'd expect those adult changes not to be biased by Input Optimization. We would also expect to see an effect on the time course of language acquisition.

Yet another interpretation of Input Optimization would be to cast it as a performance effect. The idea here is that the performance module filters the output of the grammar so as to satisfy Input Optimization. Viewing performance as a filter begs questions of teleology, but these are the same questions begged by any theory that includes constraints on the output. One might distinguish this approach from the preceding ones with psycholinguistic experiments that tap into language processing as opposed to grammatical structure per se. To the extent that we can determine different effects for the grammar and the performance system, and that Input Optimization is localized to the latter, this would militate for a view like this.

Finally, one might view Input Optimization as part of the grammar itself. On this view, Input Optimization would be an output condition on the entire grammar, a general phonological sieve. This requires: i) that the phonology itself be probabilistic in nature, an approach taken in a number of corners of the field these days;[17] and ii) that the phonology be able to constrain the syntax, morphology, and lexicon of a language. This, of course, raises the same teleological questions as above, but they are again the same as any framework that includes constraints.

The data presented here do not distinguish among these choices, but hopefully it is clear what kinds of further empirical investigations might. Do we see effects of Input Optimization in acquisition? Do we see effects of Input Optimization in adult change? Can we distinguish Input Optimization in com-

[17]See, for example, Boersma (1997), Hammond (1999), Hammond (2003), Coetzee (2008), Hayes & Wilson (2008), Pater (2009), Coetzee & Pater (2011), etc.

petence vs. performance?

Let's now turn to the second question. Why does Input Optimization not go all the way, eliminating any constraint violation? There are two reasons: constraint ranking (or weighting) and the overall functionality of the system.

In a system with weighted or ranked constraints, in some cases, it may be impossible to minimize violations of one constraint without simultaneously maximizing violations of another.

(75)

/x/	A	B
y, etc.	*!	
z, etc.		*

Here we might minimize candidates like y, maximizing candidates like z. The effect would be a less complex system, but it would not be a system free of violations.

We can imagine other configurations though. Recall the hypothetical systems (35) and (36) above, repeated as (76) and (77) below. We saw how Input Optimization would militate for the second system over the first.

(76)

	Input	Output	NC	IO-Faith
a.	/on pi/	om pi		*
b.	/an ba/	am ba		*
c.	/un bo/	um bo		*
d.	/en do/	en do		
e.	/on ta/	on ta		
f.	/un ti/	un ti		
g.	/an ku/	aŋ ku		*
h.	/in ga/	iŋ ga		*
i.	/on ke/	oŋ ke		*
			0	6

(77)

	Input	Output	NC	IO-Faith
a.	/on pi/	om pi		*
b.	/an ba/	am ba		*
c.	/en do/	en do		
d.	/on ta/	on ta		
e.	/un ti/	un ti		
f.	/in di/	in di		
g.	/an ku/	aŋ ku		*
h.	/in ga/	iŋ ga		*
			0	4

The relative complexity of the first system is: $\langle 0, 6\rangle/9 = \langle 0, 0.67\rangle$; the second: $\langle 0, 4\rangle/8 = \langle 0, 0.5\rangle$. If this is so, we might well imagine that the system could go even further:

(78)

	Input	Output	NC	IO-Faith
a.	/en do/	en do		
b.	/on ta/	on ta		
c.	/un ti/	un ti		
d.	/in di/	in di		
			0	0

Here no constraints are violated, so the system is the minimum of complexity: $\langle 0, 0 \rangle$. The effect is to reduce the inventory of nasals and stops in this environment to just those that do not violate NC or IO-Faith.

But that way madness lies. A system that allowed free rein to Input Optimization is one where no constraints are violated, effectively only one word is possible, composed of maximally unmarked segments in an optimal prosodic and segmental configuration: [ta] (or something like that). The reason then that Input Optimization does not have this effect is that it is offset by the need to have a sufficiently large set of morphemes and a sufficiently large array of combinatory possibilities so as to make communication possible. Therefore as a counterforce to Input Optimization I propose the following:

(79) **Functionality**
A language must have a sufficient inventory of sounds and sufficient combinatory possibilities so as to be a reasonable vehicle for communication.

Conceptually, this does the trick. We balance Input Optimization against the functionality of the system. Clearly, however, though this captures the logic of the situation, it's still quite speculative. Turning this into something more concrete, however, requires an investigation into the morphosyntax and semantics of a language. It would also be important to put it into explicitly quantitative terms so it can be tested statistically. I leave this to further research.

8 Morphology & phonology

The RM' constraint (62) requires that we be able to distinguish morphological processes like Welsh mutation from phonological processes like English nasal assimilation.

There are a number of ways we might do this, but this seems like the least ambiguous: a phonological process is morphological in the sense intended if, in at least one context, it affects the presence or absence of the sole marker for some morphological category.

(80) Morphological process
A process is morphological if it adds or removes the sole mark of some morphological category.

Note that, on this definition, a morphological process is *not* simply one that has morphological conditioning. As we will see, a process might very well be restricted to some morphological context and not meet the bar set by (80). The definition is then not about how the process might be formalized, but about what role it plays in the morphological system. Let's go through all the cases consider thus far and show how they fit or don't fit this rubric.

First, the English cases we considered in Section 2 involving segmental and phonotactic markedness do not qualify because they are not morphologically restricted; hence they never mark some morphological category.

The English rhythm example that we treated in Section 3 also does not qualify for the same reason. It is not morphologically conditioned and thus never marks some particular morphological category. There is a different stress alternation in English that does sometimes mark morphology. The shift of stress to the left in latinate vocabulary when certain verbs undergo zero-derivation to become nouns (Chomsky & Halle, 1968; Hayes, 1981; Kiparsky, 1982).

(81)

	Verb	Noun
combat	[kəmbǽt]	[kámbæ̀t]
torment	[tòrmɛ́nt]	[tórmɛ̀nt]
transfer	[træ̀nsfŕ̩]	[trǽnsfr̩̀]
rebel	[rəbɛ́l]	[rɛ́bəl]
confound	[kənfáwnd]	[kánfàwnd]

This is a different process, however. It only affects a small set of items of Latin origin, it only applies to nouns, and it is not subject to the restriction that there must be a secondary to the left.

The Welsh mutation facts treated in Section 4 do qualify as a morphological process. Mutation is restricted to specific morphological environments and there are environments where mutation is the sole marker of some morphological category. One environment for this is after the possessive *ei* 'his/hers'. Without the optional following echoing pronoun, the sole marker of the gender difference is the mutation triggered by the possessive. In the case, of the masculine form, we have soft mutation and in the case of the feminine, we have aspirate mutation. Thus, for example, *ei mam* [i mam] can only be 'her mother' since *mam* 'mother' does not undergo mutation. Similarly, *ei fam* [i vam] can only be 'his mother' since *mam* undergoes soft mutation.

The final consonant devoicing we treated in Section 6 does not qualify as morphological on this definition. While the process is restricted to particular morphological contexts, it never occurs without some other overt marker of that morphological context. The devoicing is never the sole marker of the comparative or superlative form.

The morphological haplology cases we saw in the same section are clearly morphological. These cases involve the presence or absence of a morpheme, a morpheme that can be the sole marker of the respective morphological category, e.g. *man* vs. *man's* and *wrong* vs. *wrongly.*

Finally, the deletion of final coronal stops in English is clearly morphological in the sense intended when it deletes the past tense marker, e.g. *look* vs. *looked.*

There are, of course, other ways we might do this, but (80) is simple and captures the intuition that a process is morphological when, in at least some context, it affects whether some morphological category is expressed.

9 Conclusion

There are always alternative analyses available, and this is especially true for statistical analyses. The skewings observed above are consistent with any number of syntactic, lexical, or semantic explanation. For example, adjectives that can be made into comparatives or superlatives in Welsh could be semantically skewed. Alternatively, some of these skewings could be statistical accidents, patterns that are statistically unlikely, but arise by chance. The argument here is that we can unify all these under a single theoretical characterization, rather than a collection of unconnected explanations and appeals to chance. In addition, our account makes clear predictions about other systems, predictions not made by an approach that treats these effects as unconnected or by chance.

This proposal is certainly not out of the blue. Similar ideas have occurred in the literature. None of these have the same empirical coverage as Input Optimization.

One idea that bears some similarities is the idea that markedness correlates with number of violations (Golston, 1998; Coetzee, 2008). Input Optimization takes this several steps further by allowing applying this to faithfulness, and by allowing it to alter distributions.

The notion of using Lexicon Optimization to alter distributions is presaged in diachronic restructuring contexts by Bermúdez-Otero (1998).

The idea that the frequency of forms is governed by constraint weights is also pursued by Hayes & Wilson (2008). Their approach uses the distributions to fix the weights. The approach here uses the categorical phonology to determine the weights and then uses those weights to determine the distribution.

The idea of Input Optimization is explicitly introduced in Hammond (2013) and Hammond (2014). The former identifies the effect for phonological markedness and faithfulness; the latter first observes the challenge posed by Welsh mutation and suggests a solution using RM. In this paper, these ideas are taken further by demonstrating that the empirical contrast between mutation and the initial English cases is indeed based on the morphological nature of mutation. This is done by analyzing the English haplology examples, the Welsh stem-final devoicing examples, and deletion of [t,d] in English in the previous section. It's also demonstrated here that RM must be revised as RM', some form of ranking is necessary to accommodate the RM' examples, that PC must be assessed using some form of constraint ranking or weighting.[18]

There are, of course, questions still to answer. One question is what precisely is the nature of morphology appealed to in the RM' constraint? It's fairly clear from the long literature on mutation that it is morphological in nature. In fact,

[18]The ranking need not be strict. The same logic will work with stochastic, harmonic, noisy harmonic, or maxent weighting.

some have argued that it isn't phonology at all any more. That said, a more precise characterization of the difference between morphological processes that are subject to RM' and phonological processes that are not would be a step forward.

A second question is how much under- or over-representation should occur in relevant cases. This paper assumes that a significant difference in distributions is what Input Optimization predicts, but this just establishes a lower bound. The working hypothesis is that under- and over-representation are bounded by other modules of the grammar, that the system will under- or over-represent in conformity with Input Optimization up to the limits imposed outside the system.

For example, we've seen that constructions like *i afal* [i aval] 'to an apple' are under-represented compared to constructions like *i gath* [i gaːθ] 'to a cat'. Crudely speaking, one can assume that this under-representation is bounded by the need to have vowel-initial words for things like apples (size of vocabulary and what phonological contrasts are available), and the need to talk about apples (what kinds of circumlocutions are available). These other aspects of the larger phonological and linguistic system are well beyond the scope of this paper, but are an obvious place to look in the future.

Appendix: statistics

The principal statistical tool used here is χ^2 (Pearson's chi-square test). What it allows us to do is test whether some distribution of items is significantly different from what's expected. For example, imagine we have a fair coin and throw it 10 times and it comes up heads 7 times. Is the coin fair? Here the expectation is that we'd get heads half the time, but in this instance we get somewhat more than that. For a linguistic example, imagine we expect words beginning with labials to occur just as often as words beginning with dorsals. In some sample of speech or text, we find 40 words beginning with labials and 60 words beginning with dorsals. Is this distribution significantly different from what we expect?

The χ^2 value can be calculated straightforwardly:

(82) $$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Here, O_i is the observed value for some cell and E_i is the expected value for that cell. In the coin example above, there are two cells: heads and tails. Given that it is a fair coin and we throw it 10 times, we expect 5 in each cell. We then get the following math:

(83) $$\frac{(7-5)^2}{5} + \frac{(3-5)^2}{5} = \frac{4}{5} + \frac{4}{5} = 1.6$$

For the linguistic example above, we expect 50 in each cell, and we have this math:

(84) $$\frac{(40-50)^2}{50} + \frac{(60-50)^2}{50} = \frac{100}{50} + \frac{100}{50} = 4$$

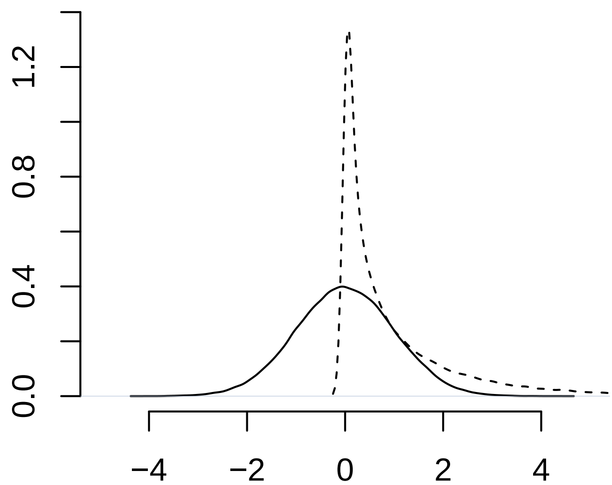

Figure 1: χ^2 distribution superimposed on a normal distribution

These χ^2 values are compared against the χ^2 distribution to determine if these departures from what's expected are significant. The χ^2 distribution is defined in terms of a normal distribution. A χ^2 distribution for one degree of freedom, like the examples above, is simply the squared normal. See Figure 1. Specifically, to determine whether a χ^2 value indicates a distribution significantly different from what's expected, we compare the χ^2 value to the distribution to see what percent of the distribution falls to the right of the value. If that percent is less than .05 of the total, then the distribution is significantly different from expected.

For the coin example, the distribution is not significantly different from expected: $X^2(1, N = 10) = 1.600$, $p = 0.206$. For the linguistic example, however, the distribution is significantly different from expected: $X^2(1, N = 100) = 4.000$, $p = 0.046$.

In most of the examples to follow, I will be explicitly comparing two distributions. For example, we might have the distribution for words beginning with dorsals vs. words beginning with labials above and wish to know whether it is significantly different from a distribution of 10 word-final labials and 12 word-final dorsals. In these cases, I define one of the distributions as the expected distribution and test whether the other is significantly different.[19]

In these cases, I will always define the distribution with the fewer tokens as the distribution to be tested and define the expected probabilities in terms of the distribution with more tokens. In the case at hand, this means we would test the observed distribution of 10 and 12 against the expected proportions of .4 and .6: $X^2(1, N = 22) = 0.273$, $p = 0.602$.[20] This is a more stringent test than doing it the other way around, testing 40 and 60 against .45 and .55: $X^2(1, N = 100) = 1.200$, $p = 0.273$.

Consider now the data treated in (7) and (8). The first comparison given in (8) is between non-premoninal ($6785 + 970 + 118 + 115 = 7988$) and prenominal tokens ($11136 + 950 + 304 + 140 = 12530$). The second prenominal distribution has more tokens, so it is the distribution to be tested against: 0.889, 0.0758, 0.0243, 0.0112. Expected values for the non-prenominal distribution are calculated by multiplying these values by the total for the non-prenominal distribution:

(85)	$\acute{\sigma}\breve{\sigma}$	0.888747	$\times$	7988	$=$	7099.311
	$\breve{\sigma}\acute{\sigma}$	0.07581804	$\times$	7988	$=$	605.6345
	$\acute{\sigma}\grave{\sigma}$	0.02426177	$\times$	7988	$=$	193.803
	$\grave{\sigma}\acute{\sigma}$	0.01117318	$\times$	7988	$=$	89.2514

Results of a χ^2 are presented in standard APA format as above. However, following linguistic practice, in addition to observed values (O), I'll also give expected values (E) and observed/expected ratios (O/E).

References

Anderson, Skye (2015). The distribution of phonological changes in Welsh plurals. U. of Arizona manuscript.

Berkley, Deborah Milam (1994a). The OCP and gradient data. *Studies in the Linguistic Sciences* **24**:59–72.

Berkley, Deborah Milam (1994b). Variability and Obligatory Contour Principle effects. In Katherine 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: Chicago Linguistic Society, 1–12.

Bermúdez-Otero, Ricardo (1998). Prosodic optimization: the Middle English length adjustment. *English Language and Linguistics* **2**:169–197.

[19] Another possibility is to test whether the entire four-way distribution departs from what's expected, but this asks a different question, whether the dimensions that define the grid are independent, and thus examines the specific values across all four cells.

[20] To calculate the expected proportions here, we simply divide the occurring values by the

Boersma, Paul (1997). How we learn variation, optionality, and probability. ROA #221.

Bolinger, D. (1962). Binomials and pitch accent. *Lingua* **11**:34–44.

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

Coetzee, Andries & Joe Pater (2011). The place of variation in phonological theory. In John Goldsmith, Jason Riggle & Alan Yu (eds.), *The Handbook of Phonological Theory*. Oxford: Blackwell, 401–434. ROA #946.

Coetzee, Andries W. (2008). Grammaticality and ungrammaticality in phonology. *Language* **84**:218–257.

Coetzee, Andries W. & Shigeto Kawahara (2013). Frequency biases in phonological variation. *Natural Language and Linguistic Theory* **31**:47–89.

Deuchar, M., P. Davies, J.R. Herring, M.C. Parafita Couto & D. Carter (2014). Building bilingual corpora: Welsh-English, Spanish-English and Spanish-Welsh. In E.M. Thomas & I. Mennen (eds.), *Advances in the Study of Bilingualism*. Bristol: Multiningual Matters.

Ellis, N. C., C. O'Dochartaigh, W. Hicks, M. Morgan & N. Laporte (2001). Cronfa electroneg o Gymraeg (CEG): A 1 million word lexical database and frequency count for Welsh. `http://www.bangor.ac.uk/canolfanbedwyr/ceg.php.en`.

Frisch, S., N. R. Large & D. B. Pisoni (2000). Perception of wordlikeness: effects of segment probability and length on the processing of nonwords. *Journal of Memory and Langauge* **42**:481–496.

Frisch, Stefan (1996). *Similarity and Frequency in Phonology*. Ph.D. thesis, Northwestern University.

Golston, Chris (1998). Constraint-based metrics. *Natural Language and Linguistic Theory* **16**:719–770.

Greenberg, J.H. (1954). A quantitative approach to the morphological typology of language. In R. F. Spencer (ed.), *Method and perspective in anthropology*. Minneapolis: University of Minnesota Press, 192–220.

Greenberg, J.H. (1966). *Language universals, with specific reference to feature hierarchies*. The Hague: Mouton.

Greenberg, J.H. (1974). *Language typology: A historical and analytic overview*. The Hague: Mouton.

Guy, Gregory R. (1991). Explanation in a variable phonology: An exponential model of morphological constraints. *Language Variation and Change* **3**:1–22.

total: $.4 = \frac{40}{40+60}$ and $.6 = \frac{60}{40+60}$.

Hammond, Michael (1988). *Constraining Metrical Theory: A Modular Theory of Rhythm and Destressing.* New York: Garland. 1984 UCLA doctoral dissertation.

Hammond, Michael (1999). Lexical frequency and rhythm. In Michael Darnell et al. (eds.), *Functionalism and Formalism in Linguistics.* Amsterdam: John Benjamins, 329–358.

Hammond, Michael (2003). Phonotactics and probabilistic ranking. In A. Carnie, H. Harley & M. Willie (eds.), *Formal Approaches to Function in Grammar: In Honor of Eloise Jelinek.* Amsterdam: Benjamins, 319–332.

Hammond, Michael (2013). Input optimization in English. *Journal of the Phonetic Society of Japan* **17**:26–38.

Hammond, Michael (2014). Phonological complexity and input optimization. *Phonological Studies* **17**:85–94.

Hammond, Michael, Edith Moravcsik & Jessica Wirth (1989). The explanatory role of language universals. In Michael Hammond, Edith Moravcsik & Jessica Wirth (eds.), *Studies in Syntactic Typology.* John Benjamins, 1–22.

Hannahs, S. J. (2013). *The Phonology of Welsh.* Oxford: Oxford University Press.

Hayes, Bruce (1981). *A Metrical Theory of Stress Rules.* New York: Garland. 1980 MIT doctoral dissertation.

Hayes, Bruce (1984). The phonology of rhythm in English. *Linguistic Inquiry* **15**:33–74.

Hayes, Bruce & Colin Wilson (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* **39**:379–440.

Hinrichs, Lars & Benedikt Szmrecsanyi (2007). Recent changes in the function and frequency of Standard English genitive constructions: a multivariate analysis of tagged corpora. *English Language and Linguistics* **11**:437–474.

Jakobson, R. (1968). *Child Language, Aphasia, and Phonological Universals.* The Hague: Mouton.

King, Gareth (2003). *Modern Welsh: a Comprehensive Grammar.* London: Routledge.

Kiparsky, Paul (1982). Lexical morphology and phonology. In I.-S. Yang (ed.), *Linguistics in the Morning Calm.* Seoul: Hanshin, 3–91.

Kučera, H. & W.N. Francis (1967). *Computational Analysis of Present-day American English.* Brown University Press.

Kurisu, Kazutaka (2001). *The Phonology of Morpheme Realization.* Ph.D. thesis, University of California Santa Cruz.

Liberman, Mark & Alan Prince (1977). On stress and linguistic rhythm. *Linguistic Inquiry* **8**:249–336.

Maddieson, Ian (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press.

McCarthy, John & Alan Prince (1993). Prosodic morphology. U. Mass.

Menn, Lise & Brian MacWhinney (1984). The repeated morph constraint: Toward an explanation. *Language* **60**:519–541.

Nevins, Andrew & Bert Vaux (2007). Underlying representations that do not minimize grammatical violations. In Sylvia Blaho, Patrik Bye & Martin Krämer (eds.), *Freedom of analysis?* Berlin: Mouton de Gruyter, 35–61.

Pater, Joe (2009). Weighted constraints in generative linguistics. *Cognitive Science* **33**:999–1035.

Pitt, M.A., L. Dilley, K. Johnson, S. Kiesling, W. Raymond, E. Hume & E. Fosler-Lussier (2007). Buckeye corpus of conversational speech (2nd release). `http://www.buckeyecorpus.osu.edu`.

Prince, Alan & Paul Smolensky (1993). Optimality Theory. U. Mass and U. of Colorado.

Shih, Stephanie, Jason Grafmiller, Richard Futrell & Joan Bresnan (2012). Rhythm's role in genitive construction choice in spoken English. In Ralf Vogel & Ruben van de Vijver (eds.), *Rhythm in Phonetics, Grammar and Cognition*. Mouton. To appear.

Stemberger, J.P. (1981). Morphological haplology. *Language* **57**:791–817.

Toutanova, Kristina, Dan Klein, Christopher Manning & Yoram Singer (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of HLT-NAACL 2003*. 252–259.

Trubetzkoy, Nikolai (1939). *Grundzüge der Phonologie*.

Turton, Danielle (2012). The darkening of english /l/: A Stochastic Stratal OT analysis. `http://ling.auf.net/lingbuzz/001524`.

Ussishkin, Adam (2000). *The Emergence of Fixed Prosody*. Ph.D. thesis, University of California, Santa Cruz. Distributed by SLUG Pubs, Department of Linguistics, University of California, Santa Cruz.

Vaux, Bert & Bridget Samuels (2005). Laryngeal markedness and aspiration. *Phonology* **22**:395–436.

Wolf, Matthew (2007). For an autosegmental theory of mutation. In L. Bateman, M. O'Keefe, E. Reilly & A. Werle (eds.), *University of Massachusetts Occasional Papers in Linguistics 32: Papers in Optimality Theory III*. Amherst, MA: GLSA, 315–404.

Zamuner, Tania, LouAnn Gerken & Michael Hammond (2004). Phonotactic probabilities in young children's speech production. *Journal of Child Language* **31**:515–536.

Zwicky, Arnold M. (1987). Suppressing the Zs. *Journal of Linguistics* **23**:133–148.