August 2010

Frequency and Other Biases in Phonological Variation

Andries W. Coetzee, University of Michigan

Shigeto Kawahara, Rutgers University

Abstract. In the past two decades, variation has received a lot of attention in mainstream generative phonology, and several different models have been developed to account for variable phonological phenomena. However, all existing generative models of phonological variation model the overall rate at which some process applies in a corpus, and therefore implicitly assume that all words are affected equally by a variable process, namely at the overall rate observed in the corpus. In this paper, we show that this is not the case. Many variable phenomena are more likely to apply to frequent than infrequent words. A model that accounts perfectly for the overall rate of application of some variable process therefore does not necessarily account very well for the actual application of the process to individual words. We illustrate this with two examples, English *t/d*-deletion and Japanese geminate devoicing. We then augment one existing generative model (noisy Harmonic Grammar) to allow for the contribution of usage frequency to the application of variable phenomena. We propose that the influence of frequency is incorporated by scaling the weights of faithfulness up or down for words of different frequencies. We show that this augmented model accounts significantly better for variation than existing generative models, and consider the extension of the augmented model to other factors such as speech register, discourse situation, etc.

Table of Contents

1. Introduction

1.1. The changing prospects of variation

Although the existence of phonological variation has been acknowledged since the early years of generative phonology (Postal 1966:185; 1968:14-15), variation received relatively little attention in mainstream generative phonology during the first 25 years of the history of this field. To the extent that variation was acknowledged, it was usually relegated to the late stages of phonology or to phonetic implementation, and was hence not considered a part of the core of phonological grammar. In Lexical Phonology, for instance, it was assumed that lexical rules apply obligatorily while "postlexical rules can be optional and subject to variation" (Kaisse & Shaw 1985:6; see also Kiparsky 1985:86).

This low valuation of variation in mainstream generative phonology contrasts with how it was viewed in the Labovian variationist tradition. This research tradition, spearheaded by Labov's work in the late 1960's (Labov 1966; 1969; etc.), developed concurrently with mainstream generative phonology, but had little impact on this field. In this approach, variation was central to grammar rather than an accidental property that applied only on the edges of grammar. In fact, Labov (2004:6) claims that variation is "the central problem of linguistics".

In the past 15 years, the prospects of variation in mainstream generative phonology have changed dramatically. It now occupies a central place in the study of phonology, and to some extent dictates the architecture of phonological grammar. A clear indication of this change is how variation has been treated in handbooks of phonological theory. The first edition of the Blackwell *Handbook of Phonological Theory* (Goldsmith 1995), which reflects the situation in generative phonology at the beginning of the 1990's, does not even contain the word "variation" in its subject index. In contrast, every handbook since contains a chapter dedicated to variation (Anttila 2007; Coetzee to appear; Coetzee & Pater to appear; Guy to appear). Similarly, several articles on variation have appeared in theoretical, generatively oriented journals over the past decade (Anttila 2002a, 2006; Anttila *et al.* 2008; Boersma & Hayes 2001; Coetzee 2006; etc.).

This same period has seen the development of several versions of current generative phonological grammar intended to deal with variation. These models have all been developed in some version of a constraint-based grammar, be that classic discrete Optimality Theory (Anttila 1997, 2002a, 2006, 2007; Anttila *et al.* 2008; Coetzee 2004, 2006, 2009c; Kiparsky 1993; Reynolds 1994), stochastic Optimality Theory (Boersma 1997; Boersma & Hayes 2001), or noisy Harmonic Grammar (Coetzee 2009a; Coetzee & Pater to appear; Jesney 2007).

In fact, variation has become so important that the ability of a grammatical model to account for variation is now often used as one of the measures of the model's sufficiency. Anttila (2002b:211) claims that an adequate theory of phonology should account for the "locus of variation" (where variation is observed and where not), and the "degrees of variation" (the frequency of different variants). Using this as a measure of success, most of the models mentioned above have been very successful. All of these models have formal mechanisms that can account for the locus of variation. With the exception of Coetzee's 2004/2006-model, these models also make predictions about the degrees of variation. In fact, they have all been shown to model the frequency with which different variants are observed very well for a range of variable phenomena.

Not only has mainstream generative phonology embraced the variationist tradition's high valuation of variation, but it has also seemingly been very successful at accounting for variation. The high degree of match between observed frequency patterns and the frequency patterns predicted by the generative models are often cited as evidence of the success of these models. An optimistic observer could conclude that the problem of variation has been resolved. In this paper, however, we will argue that such a conclusion might be too hastily drawn. In the next two sections, we show that, in spite of the close match between observed and predicted frequencies, existing models do not necessarily account very successfully for variation patterns.

1.2. Grammatical overreach

One of the persistent results of the variationist research tradition is that variation is influenced by many factors, of which grammar is but one. In fact, in a 2002 review of this tradition, Bayley identified "the principle of multiple causes" as one of the four core principles of this tradition (Bayley 2002:118). Just about every study in this tradition reports, in addition to grammar, on multiple other factors that influence the application of variable phonological processes. These factors include things like speech genre (word lists, informal conversations, read speech, etc.), discourse situation, the age, sex or educational background of the speaker, etc. When mainstream generative phonology adopted the variationist valuation of variation, it did not do so completely. The mainstream approaches to variation have all been exclusively grammatical, and have paid little or no attention to the various non-grammatical factors that also influence variation.

The reason is probably twofold. On the one hand there is a long tradition in generative linguistics of eschewing all that is not part of the *grammatical* competence of the speaker/hearer. These non-grammatical factors that influence variation can be classified as *performance*, and hence as something that falls outside of linguistics proper. This is perhaps most explicitly stated in some of Chomsky's foundational writings on the generative enterprise: "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 … surely cannot constitute the actual subject matter of linguistics, if this is to be a serious discipline." (Chomsky 1965:4.) For a more recent expression of a similar view, see Newmeyer in his presidential address to the *Linguistic Society of America*: "My goal in this article is to argue in favor of the classical Saussurean position with respect to the relationship between knowledge of language and use of language. That is, the article provides evidence in support of the idea that the mental grammar contributes to language use, but that usage, frequency, and so on are not represented in the grammar itself." (Newmeyer 2003:682.)

There is also a second practical rather than principled reason for the avoidance of non-grammatical factors. Just getting phonological theory to be able to account for any aspect of variation is a difficult enough task. The generative phonological models that were current in the early 1990's had a strong emphasis on non-variable phenomena and were not suited to handle variable phenomena. The focus on only the grammatical factors that influence variation was partially a "divide and conquer" approach. Once a model of phonological grammar has been developed that can generate variation, ways of extending this model to allow for non-grammatical factors to also impact the application of variable phenomena can be investigated. In fact, Boersma and Hayes add such a promissory note in Appendix C of their 2001 paper, where they sketch a way in which their stochastic OT grammar could be augmented to account for the influence of speech register/style on variable phonological processes.

Whatever the reason for eschewing non-grammatical factors, the fact remains that existing generative models are all grammatical models that make no formal allowance for the influence of other factors. Yet, it has been shown by the variationist tradition that phonological variation is influenced by many factors in addition to grammar. If these purely grammatical models account nearly perfectly for the data, then grammar is doing more than its fair share. Rather than evidence of an adequate model, this would be evidence of too powerful a grammar.

A successful model of phonological variation will be a model with the following properties: First, it allows grammatical and non-grammatical factors to impact the application of variable processes. Secondly, it has a well-defined mechanism for integrating the contributions of these various factors. The existing models of phonological variation in the generative tradition fall short on both of these requirements. In this paper, we develop a generative model of phonological variation that satisfies both of these requirements. Before we do so, however, we will investigate in more detail one of the non-grammatical factors that impact variation, and in the process also provide evidence that the match between observed frequency patterns and those predicted by existing models of variation is less good than what at first appears to be the case.

1.3. Usage frequency as a non-grammatical influence on variation

As mentioned above, there are many different factors in addition to grammar that influence the application of variable phonological rules. In this paper, we focus on usage frequency – i.e. the observation that some variable processes apply at different rates to words that differ in frequency. Our selection of usage frequency is one of convenience – since frequency is already quantitative, it is easier to incorporate it into a quantitative model of variation. However, other non-grammatical factors can be incorporated into the model in analogous ways. See section 5.1 for some suggestions about how this might be achieved.

We also acknowledge that usage frequency would not be considered external to the grammar in all grammatical models. In fact, in several recent models of grammar, grammar can be described as structured memory encoding of frequency – see the usage-based and exemplar models of grammar, for instance (Bybee 2001, 2006, 2007; Gahl & Yu 2006, and papers therein; Pierrehumbert 2001; etc.). In the generative tradition, however, usage frequency is not encoded in the grammar – generative models cannot treat two words differently merely because they differ in their usage frequencies. In this paper, we subscribe to a basically generative approach to grammar, and we will hence treat usage frequency as external to the grammar. See also section 5.2.

Some variable phonological processes (typically lenition processes) are more likely to affect words with higher than lower usage frequency. The result is that two words that are identical in all relevant phonological properties might be treated differently by the phonological grammar. A clear illustration is presented by Bybee who reports that the schwa in frequent *memory* is more likely to delete than the schwa in the nearly identical, but infrequent, *mammary* (Hooper 1976; see also Bybee 2000:68).

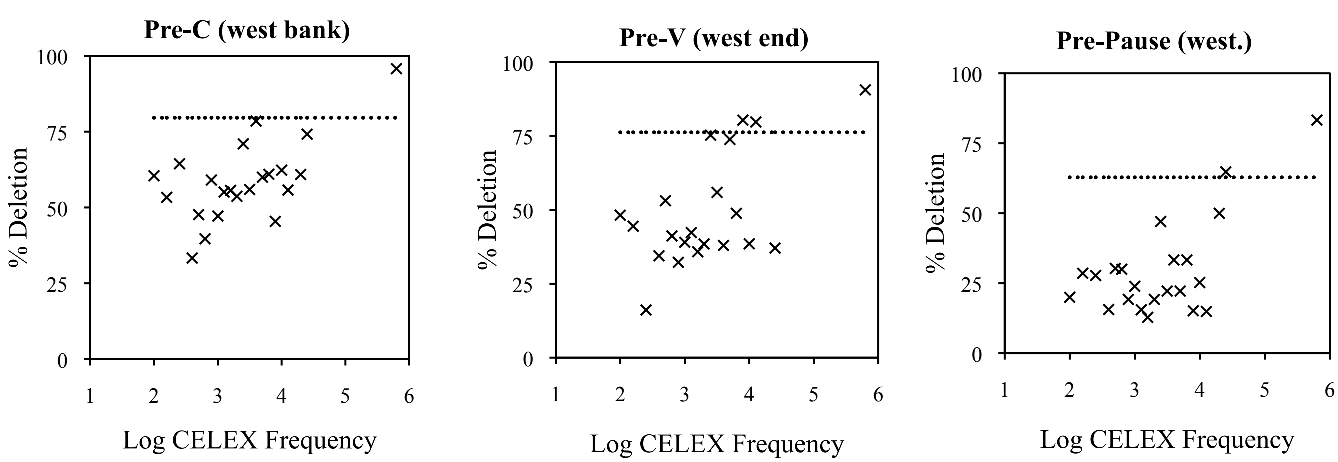
This correlation between frequency and lenition processes is widespread and has been reported for many different phonological processes. For instance, the variable deletion of word-final *t/d* from consonant clusters in English is more likely to apply to frequent than infrequent words – i.e. more deletion from frequent *just* than infrequent *jest* (Bybee 2000:69-70; 2002; Coetzee 2009a:272-273; 2009c; Lacoste 2008:187-207). The same process also applies in Dutch, and also for Dutch the correlation between frequency and the probability of deletion holds

(Goeman 1999:182; Phillips 2006:65). See section 3 for a more detailed discussion of *t*/*d*-deletion. A similar correlation of usage frequency and variation has also been illustrated for flapping in American English (Patterson & Connine 2001), word-medial *t*-deletion in English (Raymond *et al.* 2006), word-final *s*-lenition in Spanish (File-Muriel 2010), and recently for geminate devoicing in Japanese loans (on which more in section 4; see Kawahara to appear). See Phillips (2006) for a recent review of many more similar examples.

A model of variation that attempts to account for variable processes such as these in a purely grammatical manner therefore not only gives grammar more power than it is due, but also makes incorrect predictions. As we will illustrate below, the grammatical models of variation developed over the past 15 years all model the overall application of a variable process in some corpus (ignoring the identity of the words undergoing the process). The result is a model that predicts that all words will be treated the same, namely according to the overall application of the process in the corpus. The variable processes described above all apply more often to more frequent words. These words also, since they are more frequent in general, appear more often in the corpora from which data are extracted. The result is that the overall application of the variable process is usually close to the rate of application of the process in frequent words, while it is much higher than the rate of application in infrequent words. The models therefore make accurate predictions for frequent words, but inaccurate predictions for infrequent words (they predict much higher rates of application for infrequent words than are actually observed).

This problem will be illustrated in detail for English *t*/*d*-deletion (section 3) and Japanese geminate devoicing (section 4). For the sake of making the problem more concrete, we include Figure 1 below. This figure represents the rate of *t*/*d*-deletion for a selection of words from the *Buckeye Corpus* (Pitt et al. 2007), plotted against the log frequency of the words, as measured in *CELEX* (Baayen et al. 1995). (See section 3.1 on the details of how these data were extracted from the *Buckeye Corpus*.) The three panels show the rate of deletion before consonant-initial words (*west bank*), vowel-initial words (*west end*), and before pause (*west.*). The broken horizontal lines show the overall deletion rate in each context. This overall rate is close to the high rate of deletion associated with frequent words, but far above the lower rate associated with infrequent words. A model that can perfectly account for the overall deletion rate therefore does not really account very well for the data at all. The close fit in predicted deletion (the same for all words) and observed (overall) deletion achievable in the recent generative models therefore does not necessarily imply a close fit between these models and the actual data. In the rest of this paper, we develop a generative model of variation that overcomes this problem.

Figure 1: Relation between deletion rate and frequency in the *Buckeye Corpus.*



Although we will treat frequency in this paper as if it is a standalone property of a word, it is actually only one subpart of the larger concept of predictability. A word's predictability depends on many factors in addition to its frequency, as has been documented by many studies in speech processing and production over the past several decades. A word is, for instance, primed by words in its context to which it is semantically (McNamara 2005; etc.) or phonologically (Goldinger *et al.* 1992; etc.) related, or by repetition (Versace & Nevers 2003; etc.). On the other hand, a word is inhibited (i.e. becomes less predictable) if it inhabits a dense lexical neighborhood (Luce & Pisoni 1986; Vitevitch & Luce 1998; 1999; etc). Many studies have documented that factors such as these influence speech production, with the general result being that less predictable words (inhibited or less strongly primed) tend to be produced more slowly, and with more effort or clarity (Baese-Berk & Goldrick 2009; Bell *et al.* 2009; Gahl 2008; Jurafsky *et al.* 2001; Scarborough 2004; to appear; etc.). Ultimately, it would be necessary to determine an overall measure of the predictability of a word that includes contributions from all of these aspects. Our focus on usage frequency is only an initial step.

2. Noisy Harmonic Grammar with weight scaling

We develop our model in a noisy version of Harmonic Grammar (henceforth HG; Pater 2009; Smolensky & Legendre 2006). HG is a constraint-based theory that is closely related to Optimality Theory (henceforth OT; Prince & Smolensky 1993, 2004) and, in fact, an historical predecessor of OT (Goldsmith 1993; Legendre *et al.* 1990). The main difference between HG and OT is that HG works with weighted rather than ranked constraints. Noisy HG is a stochastic implementation of HG, similar to the noisy implementation of OT, known as stochastic OT (Boersma 1997; Boersma & Hayes 2001). Noisy HG and stochastic OT are in fact so closely related that we could have developed our model in this paper just as successfully in stochastic OT rather than noisy HG (see Coetzee & Pater to appear for evidence that noisy HG and stochastic OT account for most variable phenomena equally well). The main reason for opting for noisy HG over stochastic OT is based on the learnability of these two types of grammars. On the one hand, Pater (2008) has shown that there are logically possible grammars that the learning algorithm of stochastic OT cannot successfully learn. On the other hand, Boersma and Pater (2008) have proved the convergence properties of the learning algorithm of noisy HG. Although either stochastic OT or noisy HG could therefore account equally well for the data with which we work in this paper, we opt for noisy HG due to the superiority of its learning algorithm.

In the rest of this section, we first show how variation is accounted for in noisy HG, and then how we will augment this model to incorporate the influence of frequency on variation.

2.1. Noisy Harmonic Grammar

HG, like OT, is a constraint-based theory of grammar. The main difference between HG and OT is that OT relies on constraint ranking, and HG on constraint weighting. This difference is illustrated in the tableaux in (2) using the familiar OT constraints in (1). These tableaux represent the grammar of a language that does not allow tautosyllabic consonant clusters, and that repairs such clusters via deletion. In the HG tableau, *w*(Con) stands for the weight of constraint Con.

(1) MAX — Assign one violation mark for every segment in the input that lacks a correspondent in the output (no deletion). (McCarthy & Prince 1995:371.)

DEP	Assign one violation mark for every segment in the output that lacks a correspondent in the input (no epenthesis). (McCarthy & Prince 1995:371.)
*COMPLEX	Assign one violation for every tautosyllabic consonant cluster. (Prince & Smolensky 1993:96.)

(2) a. Optimality Theory: DEP >> *COMPLEX >> MAX

/lʊst/	DEP	*COMPLEX	MAX
lʊst		*!	
☞ lʊs			*
lʊs.ti	*!		

b. Harmonic Grammar: w(DEP) > w(*COMPLEX) > w(MAX)

/lʊst/	5 DEP	1.5 *COMPLEX	1 MAX	H
lʊst		-1		-1.5
☞ lʊs			-1	-1
lʊs.ti	-1			-5

In HG, each constraint is weighted, and these weights are indicated with Arabic numerals above the constraint names in HG tableaux. Constraint violations are marked with negative whole numbers rather than asterisks. A *harmony score* H is calculated for every candidate, using the formula in (3) – i.e. by taking the product of the weight of each constraint and the violation index of the candidate in terms of the constraint, and summing these products. These H-scores are indicated in the last column of the tableau. The H-score of the first candidate, for instance, is calculated as follows: The weight of DEP (5) is multiplied by the violation index of the candidate in terms of DEP (zero, since this candidate does not violate DEP). The weight of *COMPLEX (1.5) is then multiplied with the violation index of the candidate for *COMPLEX (-1), giving -1.5. Similarly, the weight of MAX (1) is multiplied with the violation index of the candidate (zero again). Finally, these products are summed, giving an H-score of -1.5 for this candidate. The candidate with the highest H-score is selected as output, and since H-scores are negative, this is the candidate with the H-score closest to zero.

(3) $$\mathsf{H}(cand) = \sum_{i=1}^{n} w_i C_i(cand)$$

Where w_i is the weight of constraint C_i, and $C_i(cand)$ is the number of violations of candidate *cand* in terms of C_i expressed as a negative integer.

The version of HG illustrated above is not noisy HG, and cannot generate variation – given these constraints and these weights, the grammar will always map the input /lʊst/ onto the output [lʊs]. However, HG has an implementation known as "noisy HG" that can generate variable outputs (Coetzee 2009a; Coetzee & Pater to appear; Jesney 2007). Noisy HG is closely related to stochastic OT (Boersma 1997; Boersma & Hayes 2001). In stochastic OT, constraint ranking is along a continuous scale, rather than discrete scale as in classic OT. Every time that the grammar is used, the ranking of each constraint is perturbed by a random negative or positive noise value. Because of this noisy evaluation, the relative ranking between two constraints can differ from one occasion to the next, resulting in variation. Noisy HG shares with stochastic OT this noisy evaluation procedure. In noisy HG, a noise value, randomly selected from a normal distribution with a mean of zero, is added to the weight of each constraint each time that the grammar is used. If the weights of two conflicting constraints are close enough, this can result in their relative weights flipping around between evaluation occasions, potentially causing variation.

In (4), the HG tableau from (2) is repeated, this time with noise. In these tableaux, *w* stands for the weight of a constraint and *nz* for the noise added to a constraint at the specific evaluation occasion. The effective weight of constraints (the sum of *w* and *nz*) is given in brackets after the constraint names. In the first tableau, the weight of *COMPLEX is adjusted down by the addition of noise at -0.4, and the weight of MAX is adjusted upward by a positive noise value of 0.2. The effect is that violation of *COMPLEX is now less serious than the violation of MAX, so that the faithful candidate has the highest H-score, and is selected as output. In the second tableau, the weight of *COMPLEX is adjusted upward and that of MAX downward, so that the deletion candidate has the highest H-score and is selected as output. This shows how the same grammar (the same constraints with the same weights) can select different outputs at different evaluation occasions because of the addition of noise to the evaluation.

(4) a. Faithful candidate optimal

	w *nz* 5 -0.7	*w* *nz* 1.5 -0.4	*w* *nz* 1 0.2	
/lʊst/	DEP (4.3)	*COMPLEX (1.1)	MAX (1.2)	H
☞ lʊst		-1		-1.1
lʊs			-1	-1.2
lʊs.ti	-1			-4.3

b. Deletion candidate optimal

	w *nz* 5 -0.7	*w* *nz* 1.5 0.1	*w* *nz* 1 -0.1	
/lʊst/	DEP (4.3)	*COMPLEX (1.6)	MAX (0.9)	H
lʊst		-1		-1.6
☞ lʊs			-1	-0.9
lʊs.ti	-1			-4.3

The formula used to calculate H-scores, first given in (3), needs to be updated to allow for the influence of noise. The updated version is given in (5).

(5) $\mathsf{H}(cand) = \sum_{i=1}^{n}\left(w_i + nz_i\right) C_i\left(cand\right)$

Where w_i is the weight of constraint C_i, nz_i the noise associated with constraint C_i at this evaluation occasion, and $C_i(cand)$ is the number of violations of candidate *cand* in terms of C_i expressed as a negative integer.

The abilities of this model of phonological variation to account for a variety of variable phenomena has been illustrated inter alia by Jesney (2007), Coetzee (2009a) and Coetzee and Pater (to appear). Coetzee and Pater (to appear), in fact, show that it performs at least as well as stochastic OT. This model, however, is still an exclusively grammatical model that treats all words exactly the same. There is no place in the formula in (5) where any factor such as usage frequency can impact the H-score of a candidate. In the next section, we introduce an augmentation to this model to remedy this shortcoming.

2.2. Weight scaling

The observation to explain is that words that are used more frequently are more likely to be treated unfaithfully. This can be captured by scaling the weight of faithfulness constraints down for frequent words and up for infrequent words. Violating a faithfulness constraint will then contribute less to the H-score of a frequent word, resulting in unfaithfulness being more likely, while it will contribute more to the H-score of an infrequent word, resulting in faithfulness being more likely. This is not a novel idea. Van Oostendorp (1997) suggested that the higher likelihood of faithfulness in more formal speech registers can be captured by ranking faithfulness constraints higher in formal speech situations – an idea that echoes the concept of "carefulness weights" in Lindblom's H&H theory of speech production (Lindblom 1990). Boersma and Hayes (2001:Appendix C) similarly suggest scaling the ranking values of constraints to account for different rates of unfaithfulness observed with different speech registers.

The effect of adding such weight scaling to the model is that two words that differ in usage frequency may be evaluated differently in exactly the same grammatical context. Continuing with the example from the previous section, assume that /lʊst/ and /nʊst/ are two different words that differ in frequency such that /lʊst/ is frequent and /nʊst/ infrequent. For the sake of the illustration, assume that /lʊst/ will be associated with a weight scaling factor of -1, and /nʊst/ with a factor of +1. The weight of faithfulness constraints will be scaled down by one unit in the evaluation of /lʊst/, and up by one in the evaluation of /nʊst/. The tableaux in (6) show how this affects the evaluation of these words. In these tableaux, the same grammatical settings are used (the same constraint weights and noise values). All that differs is the scaling factors associated with the faithfulness constraints (marked by *sf* in the tableaux). The result is that frequent /lʊst/ is mapped onto its unfaithful candidate [lʊs], while infrequent /nʊst/ is mapped onto its faithful candidate [nʊst]. An updated version of the H-score formula that incorporates the scaling factor is given in (7).

(6) a. Evaluating frequent /lʊst/, with *sf* = -1

	w *nz* *sf*	*w* *nz*	*w* *nz* *sf*	
	5 0.7 -1	1.5 0.1	1 0.2 -1	
/lʊst/	DEP (4.7)	*COMPLEX (1.6)	MAX (0.2)	H
lʊst		-1		-1.6
☞ lʊs			-1	-0.2
lʊs.ti	-1			-4.7

b. Evaluating infrequent /nʊst/, with *sf* = +1

	w *nz* *sf*	*w* *nz*	*w* *nz* *sf*	
	5 0.7 1	1.5 0.1	1 0.2 1	
/nʊst/	DEP (6.7)	*COMPLEX (1.6)	MAX (2.2)	H
☞ nʊst		-1		-1.6
nʊs			-1	-2.2
nʊs.ti	-1			-6.7

(7) $$\mathsf{H}(cand) = \sum_{i=1}^{n} \left(w_i + nz_i\right) M_i\left(cand\right) + \sum_{j=1}^{m} \left(w_j + nz_j + sf\right) F_j\left(cand\right)$$

Where M_i is the i-th markedness constraint, w_i the weight associated with M_i, nz_i the noise associated with M_i at this evaluation occasion, and $M_i(cand)$ the number of violations of candidate *cand* in terms of M_i expressed as a negative integer; and where F_j is the j-th faithfulness constraint, w_j the weight associated with F_j, nz_j the noise associated with F_j at this evaluation occasion, and $F_j(cand)$ the number of violations of candidate *cand* in terms of F_j expressed as a negative integer; and where *sf* is the scaling factor associated with the specific word being evaluated.

The same effect that we achieve here by scaling faithfulness weights could also be achieved by scaling markedness weights, or even by scaling the weights of both markedness and faithfulness constraints. In fact, Boersma and Hayes (2001:Appendix C) propose scaling the ranking values of both markedness and faithfulness constraints to incorporate style effects into their stochastic OT model. Although there are subtle differences in the patterns for which these different options can account, any of these options could have accounted equally well for the data that we discuss in this paper. We leave the question of the difference between these options for future research.

2.3. A linking function between frequency and scaling factors

The final part of our model that still needs to be specified is a linking function between frequency and scaling factors: Given a word of some frequency, what is the scaling factor that should be used in evaluating this word? This problem could be approached from two different directions. One possibility is that the mapping between frequency and scaling factors has to be learned on a language-by-language basis. The language learner will then have to take note of how words that are equivalent in their phonological properties but differ in frequency are treated differently by the grammar. From this information, he/she will deduce a function that best maps from frequency to scaling factors. Since the linking function is then determined fully by the properties of the language experience, it follows that there is no need for different languages to pattern similarly – we would not necessarily expect to see universal tendencies in how frequency maps to scaling factors. See Coetzee (2009a) for an implementation of this kind of approach.

A different possibility is that the mapping from frequency to scaling factors is at least partially independent from individual languages – i.e., that there is some universal linking function that applies similarly to all languages. The expectation would then be that frequency has the same basic influence in all languages, so that mapping from frequency to scaling factors should show the same basic properties in all languages.

Unfortunately, there is not currently enough data available on how frequency and variation interact to choose definitively between these two options. In the absence of enough such data, we explore the stronger hypothesis, namely that the same basic linking function applies in all languages. In this paper, we illustrate how such a universal mechanism accounts well for two different variable phenomena in two unrelated languages (*t*/*d*-deletion in English, and geminate devoicing in Japanese). We cautiously interpret the success of our approach in accounting for these two phenomena as evidence in favor of the approach that we take.

We propose that every word is associated with a distribution function, whose shape is determined by the frequency of the word. These functions are modeled as instantiations of the *beta* distribution (Gupta & Nadarajah 2004), and the scaling factor associated with a word is read off its distribution function. The formula of the *beta* distribution is given in (8). In addition to its argument *x*, the distribution has three parameters. ρ specifies the range of the function as spanning from -ρ to ρ. α and β are shape parameters that determine the skewness of the distribution. When α = β, the distribution is symmetric around zero. When α > β, it is left-skewed, and when α < β, it is right skewed. Additionally, the larger the difference between α and β, the more severe the skewness of the distribution is.

(8) $$f(x,\alpha,\beta,\rho) = \rho \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 x^{\alpha-1}(1-x)dx}$$

Frequent words must have a negative scaling factor, and infrequent words a positive one. But what counts as “frequent” or “infrequent”? A reference frequency has to be established such that words that appear more frequently than this will be treated as frequent, and words that appear less frequently as infrequent. There are several ways in which such a reference point can be established. The average or median frequency of all the words in the lexicon could be used, for instance. We explored several different options, and settled on the one that resulted in the best fit of our model to the data. Specifically, we settled on a decision that results in (at least) half of the tokens in the corpus being treated as frequent, and (at most) half being treated as infrequent. The exact way in which we determine the reference frequency is stated in (9).

(9) Let *N* be the total number of tokens in the corpus.

i. Order the words in the corpus in terms of frequency.

ii. Determine the point on this ordering such that at least *N*/2 of all the tokens are above this point.

iii. Determine the log frequency of the word just above this point, and the word just below this point.

iv. Let the reference frequency be halfway between these two log frequencies.

We illustrate how this works with an example. In section 3, we work with a corpus on *t*/*d*-deletion, extracted from the *Buckeye Corpus* (Pitt *et al.* 2007). Our corpus contains 16,460 tokens. Ordering the tokens according to their CELEX frequencies (Baayen *et al.* 1995), the word *and* occupies the topmost position. It also accounts for more than half of the tokens in the corpus (*and* appears 8,827 times in our corpus). The reference point that we use is halfway between the log CELEX frequency of *and*, and the log CELEX frequency of *just*, the next most frequent word in our corpus. For reasons that we explain in section 3, we grouped words together into larger log groups. *just* was placed into the 4.4 log frequency group, while *and* was placed into the 5.8 log frequency group. The midpoint between these two is 5.1, and this is the reference frequency that we used in our modeling of the data in our *t*/*d*-deletion corpus.

Having established the reference frequency, the values of the shape parameters (α and β) of the *beta* distribution associated with each word, as well as the scaling factor associated with each word, can now be determined. Specifically, we propose that α is set equal to the log reference frequency, and β to the log frequency of the specific word. Abstracting away from the contribution of ρ for the moment, the scaling factor associated with the word is then the mode of the specific instantiation of the *beta* distribution that results. As shown in (10), the result is that scaling factors of frequent words will be negative, and those of infrequent words will be positive.

(10) Determining the values of α, β, and the scaling factor associated with each word

Frequency	Shape parameters	Skewness	Mode (= scaling factor)
< reference	$\alpha > \beta$	Left	positive
= reference	$\alpha = \beta$	Symmetrical	zero
> reference	$\alpha < \beta$	Right	negative

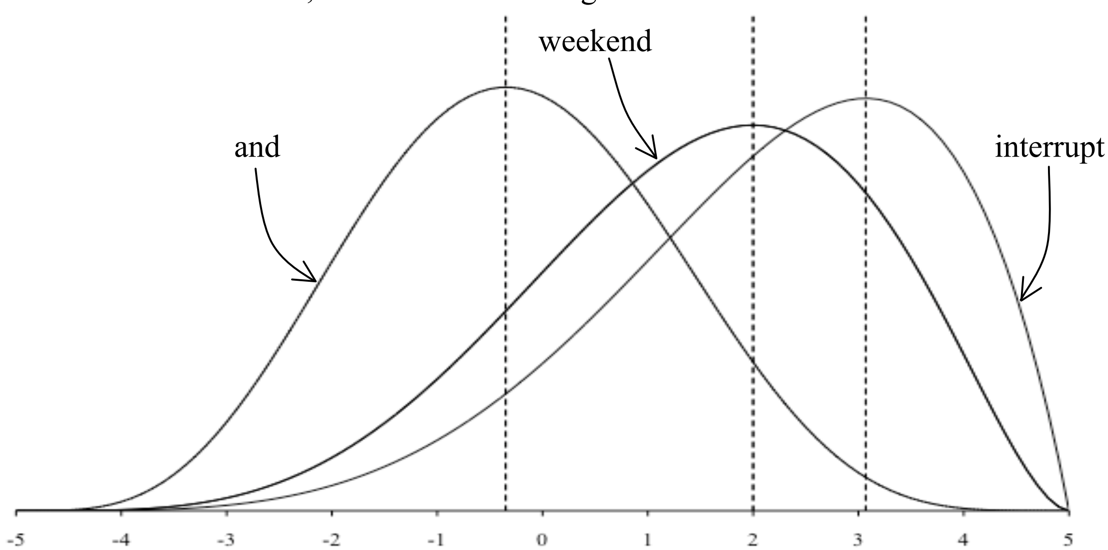
The last parameter to account for is the range parameter ρ. ρ does not influence the shape of the *beta* distribution, but only its range. In particular, it specifies the minimum and maximum value of the function on the *x*-axis. Since ρ does not influence the shape of the distribution, it also does not influence the sign of the mode of the distribution, but only its absolute value – the higher the value of ρ, the higher the absolute value of the mode. The higher ρ is, the higher the scaling factors will be. And the higher the scaling factors, the more influence the frequency of words can have on their evaluation. ρ therefore determines how much frequency is allowed to influence how the grammar functions. We propose that the value of ρ be fit to the data – i.e. for every corpus, the value of ρ that results in the best fit between the model and the data is used.

In (11) we give examples of the parameter values and the modes for three words from our *t*/*d*-deletion corpus. *and* is used as an example of a frequent word. *and*'s distribution function is right skewed, so that the mode of this function, and hence *and*'s scaling factor, is negative. *interrupt* and *weekend* both appear less frequently than the reference frequency, and both serve as examples of infrequent words. Their distributions are left skewed, so that their modes are positive, and the scaling factors associated with these two words are also positive. Although both *interrupt* and *weekend* are infrequent, they differ in frequency. *interrupt* has a CELEX log frequency of 1.98 and *weekend* of 2.76. In the distribution function associated with *interrupt*, the difference between the values of α and β is hence larger than in *weekend* (α = 5.1, β = 1.98 vs. α = 5.1, β = 2.76). We include both of these words to show what effect this has – the larger the difference between α and β, the more skewed the distribution, and hence the more extreme the mode of the distribution. A less frequent word therefore has a higher scaling factor. We also give the modes for these distributions at three different values of ρ. Note how a change in ρ influences only the absolute value of the modes, and not their sign. In Figure 2 we show the shape of the distribution functions for these tokens when ρ = 5 (the value that we use for ρ in section 3).[1]

(11) Examples of scaling factors in the *t*/*d*-deletion corpus (see section 3.3)

	Shape parameters			Mode (scaling factor)		
Word	α (reference frequency)	β (log frequency of word)	Skew	ρ = 1	ρ = 5	ρ = 10
and	5.1	5.71	Right	-0.07	-0.35	-0.70
weekend	5.1	2.76	Left	0.40	2.00	4.00
interrupt	5.1	1.98	Left	0.61	3.07	6.14

Figure 2: *Beta* distributions for words from (11) with ρ = 5. Vertical broken lines mark the modes for the distributions, and hence the scaling factors associated with these words.



In principle, scaling factors could be deduced from a more well-known distribution such as the normal distribution. For instance, scaling factors for frequent words could be selected from a normal distribution with a negative mean, and those for infrequent words from a normal distribution with a positive mean. Our selection of the *beta* rather than the normal distribution is motivated by the fact that the *beta* distribution has a finite range (specified by ρ), while the normal distribution has an infinite range. The finite range of the *beta* distribution places an absolute limit on the influence that non-grammatical factors can have via weight scaling. If

[1] An Excel file for the calculation of the *beta* distribution's mode under different settings of the three parameters is available from http://www.quantitativeskills.com/sisa/rojo/distribs.htm. In this file, the range parameter ρ is represented by *A* and *B*, with *A* = -ρ and *B* = ρ. The shape parameter α is represented by *p*, and β by *q*.

scaling factors were taken from the normal distribution with its infinite range, there would be no principled limit on how much non-grammatical factors could influence the application of variation. See sections 5.1 and 5.2 for more.

3. English t/d-deletion

Word-final *t/d* variably deletes from consonant clusters in English, so that a word like *west* can be grammatically pronounced as [wɛst] or [wɛs]. This deletion process is probably the most extensively studied variable phonological process. It has been described in detail for countless dialects of English (see Coetzee 2004:Chapter 5 for a review), and even for languages other than English (on Dutch, see Goeman 1999; Goeman & van Reenen 1985; Schouten 1982, 1984). Since this process has been studied so extensively, the factors (both grammatical and non-grammatical) that influence its application are reasonably well understood. We begin this section by first reviewing some of the grammatical and non-grammatical factors that are known to influence this process, focusing on those aspects for which we will provide an account. We then develop a purely grammatical account in the noisy HG framework (see 2.1). Once the grammatical account has been established, we augment it to account for the influence of usage frequency according to the method described above in 2.2 and 2.3.

3.1. Grammatical and non-grammatical influences

Rather than reviewing all of the grammatical factors that are known to influence *t/d*-deletion, we mention some of these factors as examples. The aim is to show that this process is influenced by the same kinds of grammatical considerations as those that influence "ordinary" non-variable phonological rules. Echoing a sentiment that has been present throughout the variationist research tradition for nearly 40 years, Anttila (1997:44) uses this as a motivation that phonological grammar should be expected to account for at least part of variation: "… if variation preferences are based on phonological variables, then it seems reasonable to expect phonology to make sense of them."

In a summary of the grammatical factors that influence *t/d*-deletion, Labov (1989) includes the following: (i) *Stress*. *t/d* is more likely to delete from an unstressed syllable (*cubist*) than a stressed syllable (*insist*). (ii) *Cluster size*. Deletion is more likely from tri-consonantal (*tanked* [tæŋkt]) than from bi-consonantal clusters (*tacked* [tækt]). (iii) *Identity of preceding segment*. Deletion is more likely after consonants that share more features with *t/d* than consonants that share fewer features – more deletion from *kissed*, where [s] shares place (coronal) and sonorancy (non-sonorant) with the following [t], than from *seemed* where [m] shares no major features with the following [d]. (iv) *Morphology*. *t/d* that functions as the past tense suffix of a regular past tense (*missed*) is less likely to delete than *t/d* that functions as the past tense suffix in a semi-weak verb (*kept*), which is less likely to delete than *t/d* that is part of a morphological root (*mist*).

Another grammatical factor that influences *t/d*-deletion is the context that follows the word-final *t/d*. We will use this factor as an example of a grammatical factor in the rest of this section, and therefore discuss it in more detail. In every dialect of English for which *t/d*-deletion has been studied, it has been found that deletion is most likely if the next word begins with a consonant (*west bank*). Dialects diverge on whether a following vowel-initial word (*west end*) or a pause (*west.*) results in more deletion. The table in (12) contains a sample of the data available on the influence of the following context.[2] The data on all but Columbus English are taken from the literature, with references given in a footnote. See just below on how the data on Columbus English were attained.

(12) Percent *t/d*-deletion in different English dialects in pre-consonantal, pre-vocalic, and pre-pausal contexts.[3]

Relative deletion rate		Pre-V *west end*	Pre-Pause *west.*	Pre-C *west bank*
Pre-C > Pre-Pause > Pre-V	AAVE (Washington, DC)	29	73	76
	Jamaican English	63	71	85
	New York City English	66	83	100
	Tejano English	25	46	62
	Trinidadian English	21	31	81
Pre-C > Pre-V > Pre-Pause	Philadelphia English	38	12	100
	Chicano English	45	37	62
	Columbus English	76	63	80

The data on Columbus English were extracted from the *Buckeye Corpus* (Pitt *et al.* 2007). This is a corpus of conversational speech collected from 40 lifelong residents of Columbus, Ohio. All of the speech was both orthographically and phonetically transcribed. In order to compile a list of words from the corpus to which *t/d*-deletion could apply, we extracted all words that end orthographically in -*Ct* or -*Cd* (where *C* stands for any consonant). Since *t/d* that corresponds to the past tense suffix is consistently treated differently (see discussion above), and since our focus is on the influence of the phonological context, we excluded words with this suffix. The principle by which we selected tokens from the corpus already excluded past tense forms that end orthographically in -*ed*. We manually removed also the semi-weak past tense forms, such as *kept*. We also removed a few other classes of words. First, due to the difficulty of determining whether the word-final *t/d* has been realized before a word that starts with [t] or [d], we removed all such tokens from the list. Secondly, we removed words that end orthographically in *-rt/-rd* or *-lt/-ld*. These tokens showed unexpectedly low deletion rates in the corpus. We listened to a selection of these tokens, and it became clear that *r* and *l* were often realized as coloring on the preceding vowel rather than as a separate consonant, so that *-rt/-rd* and *-lt/-ld* words often do not actually end in consonant clusters phonologically. There is precedence for this observation in the variationist literature where especially *-rt/-rd* words are often found to have lower deletion rates than other *-Ct/-Cd* words (Guy & Boberg 1997). Lastly, we removed words such as *thought* and *could*, that end orthographically but not phonologically in *-Ct/-Cd*. This left a list of 16,460 tokens, representing 459 different words. The phonetic transcription in

[2] These data represent a simplification, especially with regard to the pre-consonantal context. Labov (1989) and Guy (1991), among others, show that *t/d*-deletion rates are different before consonants of different types. We follow the practice in the vast majority of the *t/d*-deletion literature of lumping all of the consonants together.

[3] Sources: AAVE (Fasold 1972), Chicano (Santa Ana 1991), Jamaican (Patrick 1992), New York City (Guy 1980), Tejano (Bayley 1995), Trinidad (Kang 1994), Philadelphia (Guy 1980). See below for how the data on Columbus English were attained.

the corpus for each of the token words was then consulted, and each token was coded as either "*t/d* deleted" or "*t/d* retained".[4] Each token was also classified as pre-consonantal, pre-vocalic, or pre-pausal based on the context in which the token appeared in the corpus.[5,6]

Several non-grammatical factors that influence the application of *t/d*-deletion have also been documented, and we again mention only some here as examples. Nearly every paper in the variationist tradition on *t/d*-deletion mentions at least some biographical factors, such as the age, sex, or ethnicity of the speaker, that influence application of the process. Additionally, speech register is also often mentioned as an influencing factor with less formal registers associated with higher deletion rates. Browman and Goldstein (1990), for instance, found little evidence of *t/d*-deletion in the reading of a word list, but they did find evidence for the process in a more casual conversational speech style. Mitterer and Ernestus (2006) studied the analogous process in Dutch in two speech corpora. One corpus consisted of read speech (literally, novels read on tape for the blind) – i.e. a rather formal speech register. The other corpus consisted of recordings of casual speech. They found evidence of deletion in both corpora, but at very different rates (8% for the read speech vs. 45% for the casual speech).

The non-grammatical factor on which we focus in this paper is usage frequency, and we therefore report on it in more detail. As we already discussed in section 1.3, reductive processes such as *t/d*-deletion usually apply at higher rates to words of higher frequency – i.e. more deletion from frequent *just* than from phonologically similar but infrequent *jest*. Bybee (2000:69-70) was the first to show that this is true for *t/d*-deletion. She reanalyzes a part of Santa Ana's 1991-corpus of Chicano English, and finds a deletion rate of 54.4% in high frequency words compared to 34.4% for low frequency words.[7] Phillips (2006:65) shows that frequency has the same influence in the analogous process in Dutch.

---

[4] A token was coded as "*t/d* deleted" only if no segment was transcribed for the underlying *t/d*. In the *Buckeye Corpus*, underlying *t/d* was transcribed with several different possible surface realizations, including faithful realizations [t] or [d], glottalized realizations [tˁ] or [dˁ], flap [ɾ], etc. All tokens transcribed with one of these realizations were coded as "*t/d* retained". Since the corpus contains no articulatory data, deletion is defined here as the absence of any acoustic evidence of *t/d*. An actually articulated *t/d* might not have any acoustic realization when it is articulated before a following labial consonant. If the labial closure of the following consonant is made before the release of the *t/d*, the potential acoustic effect of the coronal release is masked by the labial closure (Browman & Goldstein 1990). The result is that the actual articulatory *t/d*-deletion rate before consonants may be somewhat lower than the acoustic rate reported here. However, given that words that start on labials form only a small fraction of consonant-initial words of English the difference between the articulatory and acoustic deletion rates should be negligible.

[5] The coding conventions in the *Buckeye Corpus* do not actually include a category for pauses. We coded as pre-pausal the following tokens: (i) Tokens where the corpus indicates that silence followed an utterance. (ii) Tokens where the corpus indicates that an utterance was followed by the interviewer speaking, and where it was clear from the context that the interviewer did not interrupt the interviewee mid-utterance. (iii) Utterances followed by some kind of non-speech vocalization noise, and where the context made it clear that this vocalization noise did not occur mid-utterance.

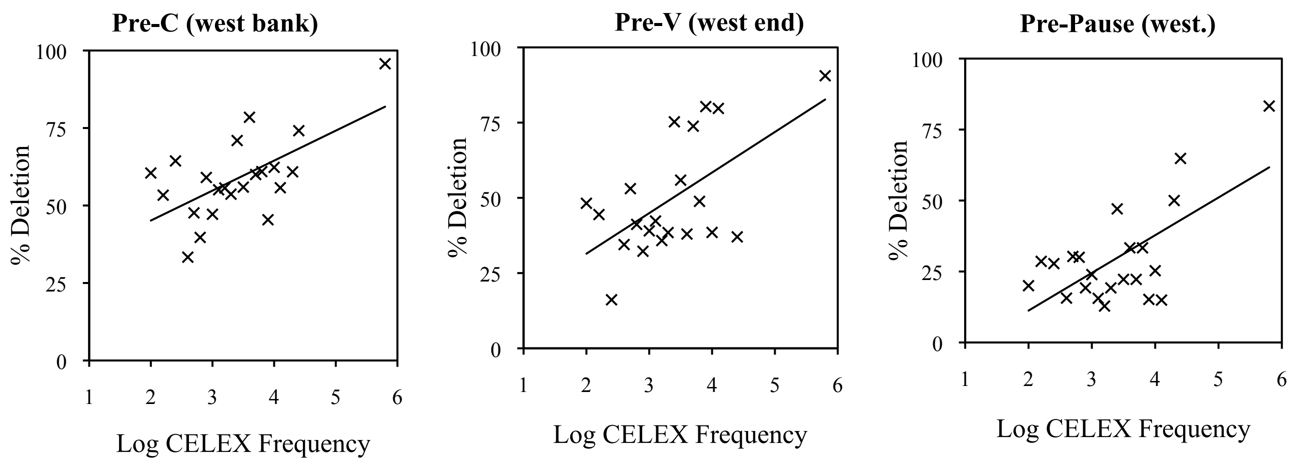
[6] The corpus of *t/d*-words that we used is included with this submission.

[7] Bybee (2001) and Jurafsky *et al.* (2001:252-255) show that mere lexical usage frequency does not capture the full influence of frequency on reductive processes. Just as important, and in some instances maybe even more important, is frequency of use within a specific syntagmatic context. That is, the [t] in *best* may delete more often from a more frequent phrase such as *best friend* than from a less frequent phrase such as *best fruit*. Although an adequate account

In order to investigate the influence of frequency on *t/d*-deletion in the *Buckeye Corpus*, we determined the frequency of each of the words that we selected from this corpus in CELEX (Baayen *et al.* 1995), and then transformed these counts by taking their logarithms (with base 10).[8,9] We then divided the words into frequency bins that span 0.1 intervals on the log-transformed frequency values. In some instances, a frequency bin contained too few tokens to calculate reliable statistics, and in those instances adjacent frequency bins were combined so that a few bins span a wider range than 0.1. In total, 23 frequency bins were created ranging in log-transformed frequency from (0 to 2.0) up to (5.7 to 5.8). The deletion rate in each of the three contexts (pre-vowel, pre-consonant, pre-pause) was then calculated for each frequency bin. This gives a data set where deletion rates in each of the contexts can be plotted against frequency to look for a correlation, as in Figure 3. This figure shows a clear positive correlation between frequency and deletion rate in all three contexts. In fact, the correlation is significant in all three contexts (Pre-C: $r^2 = .39$, $p < .001$; Pre-V: $r^2 = .36$, $p < .002$; Pre-Pause: $r^2 = .42$, $p < .001$).

In the next section, we first develop an account for the influence of the following phonological context on *t/d*-deletion in Columbus English, as given in (12). In doing this, we will abstract away from the influence of usage frequency, shown in Figure 3. Once this grammatical account is in place, we will augment it to incorporate the influence of frequency.

Figure 3: The relation between frequency and deletion rate in Columbus English in Pre-C, Pre-V and Pre-Pause contexts. The *x*-axis represents log-transformed CELEX frequencies. Deletion rate is plotted on the *y*-axis.



3.2. A grammatical account

In this section we develop a noisy HG account for the overall deletion rates observed in Columbus English, as shown in the table in (12). For similar accounts of the other data from this table, see Coetzee and Pater (to appear).

---

of phonological variation will ultimately have to incorporate all relevant types of frequency influences (and all other relevant influences), we will focus only on lexical usage frequency in this article.

8 Throughout this paper, all logarithmic transformations are done with a base of 10.

9 Since log of zero is undefined, a constant of one was added to all frequencies before they were log transformed.

3.2.1. Constraints

The constraints that we use are given in (13). The two contextual faithfulness constraints are in the spirit of Steriade's "licensing by cue" constraints – i.e., they protect segments from deletion in contexts where the cues for their perception are saliently licensed (Steriade 1999, 2001, 2008).

(13)	*CT	Assign one violation mark for every word that ends on the sequence [-Ct] or [-Cd].[10]
	MAX	Assign one violation mark for each segment in the input that does not have an output correspondent (no deletion). (McCarthy & Prince 1995:371.)
	MAX-PRE-V	Assign one violation mark for each segment that appear in pre-vocalic context in the input, and that does not have a correspondent in the output (no deletion before a vowel). (Côté 2004:22.)
	MAX-PRE-PAUSE	Assign one violation mark for each segment that appear in pre-pausal context in the input, and that does not have a correspondent in the output (no deletion before a pause).

Steriade proposes that a segment is protected by special faithfulness constraints in contexts where its perceptual cues are more robustly licensed. The consonant release burst can cue both place (Lahiri *et al.* 1984) and manner information (Stevens & Keyser 1989). The formant transitions out of a consonant also carry information about both place (Celdran & Villalba 1995; Eek & Meister 1995; Fowler 1994; Fruchter & Sussman 1997; Kewley-Port 1983; Kewley-Port *et al.* 1983; Nearey & Shammas 1987; Stevens & Blumstein 1978; Sussman *et al.* 1991; etc.) and manner (Diehl & Walsh 1989; Walsh & Diehl 1991). To motivate the existence of the positional versions of MAX, it is therefore necessary to show that release bursts and formant transitions are more robustly licensed in pre-vocalic and pre-pausal position than in pre-consonantal position.

In pre-consonantal position, the likelihood of a consonantal release being realized is relatively small (Browman & Goldstein 1990). Except when the following consonant is a sonorant, there is also no opportunity for the realization of formant transitions, and even into a following sonorant, robust transitions are less likely than into a following vowel. Pre-consonantal position is hence the context in which *t/d* is least well cued, so that there is no special faithfulness constraint that protects against deletion specifically in this context.

In pre-pausal position, formant transitions out of *t/d* cannot be realized. However, some dialects of English do release final consonants, and in such dialects the release burst is licensed pre-pausally (Holmes 1995:443). There is also evidence that utterance-final released consonants are perceived more accurately than unreleased consonants (Malécot 1958). In pre-vocalic

[10] This constraint is a special version of the more general *COMPLEX, that applies only to a subclass of consonant clusters, and only when these clusters appear in word-final position. As it stands, the constraint is too specific. For instance, deletion of [p] from words like *ramp*, *wisp*, etc., and deletion of [k] from words like *whisk*, *task*, etc. are also observed. To account for these deletions, the constraint should probably be generalized so that it penalizes all […C+stop] sequences. However, the literature contains virtually no information on the deletion of [p] and [k], probably because there are so few […Cp] and […Ck] words in English. For this reason, we assume the more specific constraint here. However, see Coetzee (2004:Chapter 5) for an exploration of a more general constraint.

position, both releases and transitions can be realized, but only across a word boundary. Although both of these contexts license perception of *t/d* more robustly than pre-consonantal context, it is not clear that there is a universal robustness difference between these contexts. The robustness difference between them in some language may depend on how likely final stops are to be released, and on how likely releases and formant transitions are to be realized across word boundaries in the language. This is also reflected in the data in the table in (12). Although all dialects show most deletion in pre-consonantal contexts, dialects differ in whether pre-vocalic or pre-pausal context shows the lowest deletion rate.

3.2.2. The learning simulation and results

The constraint weights for Columbus English were determined by running a learning simulation with *Praat*'s built-in noisy HG learning algorithm (Boersma & Weenink 2009). For details on this learning algorithm, see Boersma and Pater (2008) and Coetzee and Pater (2008). In creating an input file for the algorithm, we assumed that each of the contexts (pre-consonantal, pre-vocalic, pre-pausal) appears 100 times. Deletion was represented in the 100 tokens in each context proportional to the overall deletion rates from (12) – i.e. in pre-consonantal context, 80 tokens were coded as pronounced with deletion and 20 with a final *t/d*, etc.[11] By constructing the learning input file based on the overall deletion rate, we followed the tradition in the literature. The account that we develop here will therefore suffer from the same shortcomings as the earlier generative accounts. In the next section, we will augment our account by implementing weight scaling. In running the learning simulation, we set the "decision strategy" to "Linear OT" (*Praat*'s implementation of the noisy HG learning algorithm). All other settings were kept at *Praat*'s defaults. Once the grammar has been learned, *Praat*'s "to output distribution" function was used to test the predicted output of the grammar.

The constraint weights that were learned are given in (14). Before this grammar is used to evaluate output candidates, noise is added to the constraint weights. In the noisy HG implementation in *Praat*, this noise is randomly selected from a normal distribution with a mean of zero. Under the default *Praat* setting, the standard deviation of the distribution is set to 2. If the sum of a constraint's weight and the noise added to this weight at a particular evaluation occasion is less than zero, it is reset to zero during evaluation. This is to prevent a candidate from being rewarded in its H-score for violating a constraint – a negative constraint weight multiplied by the negative integer used to mark constraint violation will increase the H-score calculated according to the formula in (7).

(14)	*CT	101.16
	MAX	98.84
	MAX-PRE-V	-1.51
	MAX-PRE-PAUSE	0.96

In (15), we show the output patterns generated by a grammar with the weights in (14). As expected, there is a close match between the observed deletion rates (on which the learning input file was based), and the deletion rates predicted by the grammar. As has been shown before, noisy HG can replicate variation rates extremely well (Coetzee 2009a; Coetzee & Pater to appear; Jesney 2007). Given the close match between the observed and expected rate of deletion,

[11] A copy of the *Praat* input file is available with this submission.

it would appear that this model of the variation in Columbus English is very successful. However, as we have pointed out above, this is not actually the case. The deletion rate predicted by this grammar reflects the overall deletion rate in the corpus, which is heavily skewed by the deletion rates associated with frequent words. Especially for infrequent words, the predicted and observed deletion rates are actually quite far apart. This is shown in Figure 4, which plots the deletion rate predicted by the grammar against the actual deletion rates observed for words of different frequencies.

(15)

Context	Observed deletion	Expected deletion
Pre-C	80%	79.6%
Pre-V	76%	76.2%
Pre-Pause	63%	62.8%

3.3. Incorporating the frequency bias through weight scaling

To remedy the shortcomings of the purely grammatical model from the previous section, we augmented this model with weight scaling. In order to apply weight scaling, the scaling factors for words of different frequencies need to be determined, and to do that, the values of the parameters (α, β, and ρ) of the *beta* distribution associated with words of different frequencies need to be determined. We start by showing how the values of α and β are determined. As explained in section 2.3, the value of α is set to the logarithm of the reference frequency – i.e. that frequency that divides the words into frequent and infrequent sets. The method for determining this reference frequency was given in (9) in section 2.3, where it was also explained how the log reference frequency for our Columbus English *t/d*-deletion corpus was determined to be 5.1. For words of all frequencies, the value of α is hence set to 5.1. The value of β is set to the log frequency of the bin to which the word belongs. For the word *and*, for instance, α is set to 5.1, and β to the log bin to which *and* belongs, namely 5.8.

As shown in section 2.3, ρ only influences the size of the scaling factors and not their signs. Its role is to determine how much influence usage frequency (via weight scaling) can have on the functioning of the grammar. We propose that the value of ρ is determined by fitting the model to the data. This value therefore has to be determined separately for each language (represented by some corpus). To determine the value of ρ that results in the best fit to our data, we ran multiple simulations, keeping the values of α and β constant while increasing the value of ρ by whole number steps from 1 upwards. We then compared the weight scaled models with the baseline model without weight scaling in terms of their mean square errors relative to the observed deletion rates. The improvement of the weight scaled grammars at different integer values of ρ can then be compared, and the value of ρ can be selected where the improvement reaches its maximum.

In (16), we give the scaling factors associated with words belonging to different frequency bins in our corpus at different values of ρ. As the frequency increases (top to bottom), the scaling factors decrease, corresponding to the fact that faithfulness constraints play a less important role in the evaluation of more frequent words. For the most frequent frequency bin (5.8), the scaling factor is negative, since for words in this bin α (the reference value, 5.1) is smaller than β (the log frequency of the bin, 5.8), resulting in a right skewed *beta* distribution with a negative mode.

As the value of ρ increases (from left to right), the absolute values of all the scaling factors increase, even though their signs do not change. This corresponds to the fact that frequency has a larger influence (via the scaling factors) at larger values for ρ. The "baseline" column represents the basic grammar without frequency scaling.

(16) Scaling factors for words of different frequencies, at different values of ρ.

		ρ						
		Baseline	3	4	5	6	7	
Frequency bins	2.0	0.0	1.82	2.43	3.04	3.65	4.26	Scaling factor
	2.6	0.0	1.32	1.76	2.20	2.63	3.07	
	3.0	0.0	1.03	1.38	1.72	2.06	2.41	
	3.5	0.0	0.73	0.97	1.21	1.45	1.69	
	4.0	0.0	0.47	0.62	0.78	0.93	1.09	
	4.4	0.0	0.28	0.37	0.47	0.56	0.65	
	5.8	0.0	-0.24	-0.32	-0.40	-0.47	-0.55	

In (17) we show the deletion rates in pre-consonantal position predicted for a selection of frequency bins, and the values of ρ from (16). Since frequency has no influence in the baseline grammar, words of all frequencies are treated the same, and the same deletion rate is expected for all frequency bins. For all of the other values of ρ, deletion rates increase as frequency increases (top to bottom). Given that the scaling factors decrease as frequency increases, as shown in (16), this is the expected result. Lower scaling factors imply lower effective weights for faithfulness constraints, and hence higher rates of unfaithfulness. For all but frequency bin 5.8, deletion rates decrease as the value of ρ increases (left to right). These frequency bins represent words that appear less often than the reference frequency, and as shown in (16), these bins are therefore associated with positive scaling factors. Also shown in (16) is that the scaling factors increase with ρ. At higher values of ρ, the faithfulness constraints will hence have higher effective weights, and therefore exert more influence on the selection of the output, with the resulting higher rates of faithfulness. Frequency bin 5.8 is the only bin with a frequency higher than the reference frequency of 5.1. As shown in (16), the scaling factors associated with this bin are hence negative, and decrease as ρ increases. As a result, for this frequency bin, deletion rates increase as ρ increases. The contribution of ρ to the model should now be clear. Higher values of ρ result in an increased contribution of frequency to the selection of the output. If a word is frequent and therefore has a higher than overall deletion rate, its deletion rate will be even higher at higher values of ρ. On the other hand, if a word is infrequent and therefore has a lower than overall deletion rate, its deletion rate will be even lower at higher values of ρ.

The table in (18) compares the performance of the model at different values for ρ in terms of mean square errors. For each value of ρ, we also give the percent improvement of the model relative to the baseline model. The performance of the model steadily increases up to a value of 5 for ρ, after which it starts declining again. Based on this, we set the value of ρ for the Columbus English *t/d*-deletion corpus at 5.

(17) Predicted deletion rates (%) in pre-consonantal context at different values of ρ.

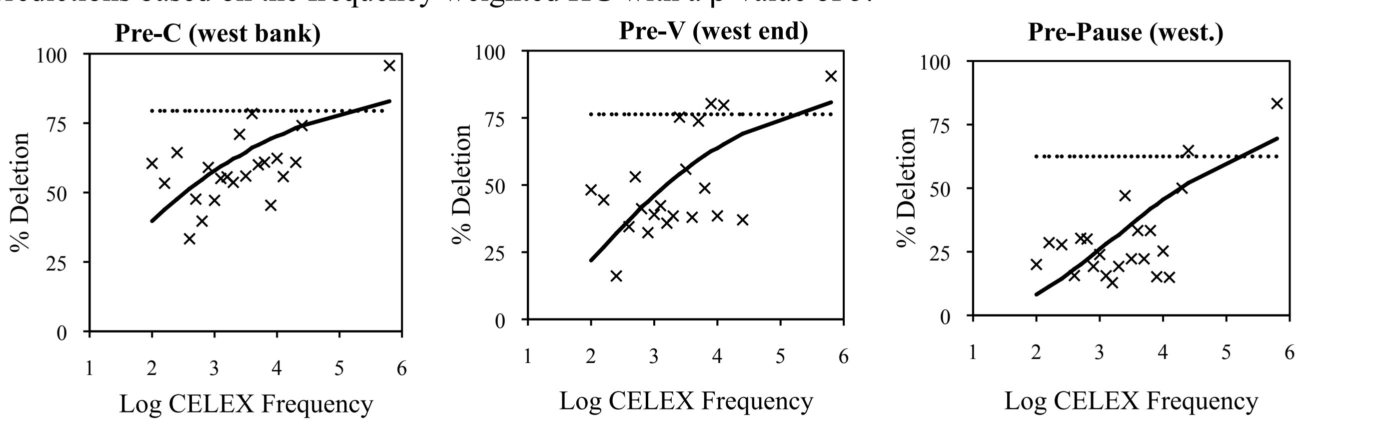
		ρ						
		Baseline	3	4	5	6	7	
Frequency bins	2.0	79.4	56.9	48.4	39.7	32.0	24.5	Predicted deletion
	2.6	79.4	63.7	58.0	51.4	45.9	39.5	
	3.0	79.4	67.7	63.0	57.9	53.6	48.9	
	3.5	79.4	71.1	68.3	64.4	62.0	58.6	
	4.0	79.4	74.5	72.8	70.3	68.5	67.1	
	4.4	79.4	76.5	75.6	74.1	73.2	72.2	
	5.8	79.4	81.6	82.4	82.9	83.9	84.2	

(18) Mean square errors and percent improvement relative to the baseline, unscaled grammar at different values of ρ.

	ρ					
	Baseline	3	4	5	6	7
Mean Square Error	1,009.7	354.0	280.4	208.2	218.8	425.9
% Improvement Relative to Baseline	0%	64.9%	72.2%	79.4%	78.3%	57.8%

Figure 4 shows the performance of the baseline model relative to a frequency scaled model with ρ = 5. The broken line represents the baseline model that treats all words the same, in terms of the overall deletion rate. The solid line represents the frequency scaled model. It predicts a higher than overall deletion rate for words in frequency bin 5.8, and lower than overall deletion rates for other frequency bins. This figure also shows that the frequency scaled model fits the data better than the baseline model. In fact, as shown in (18), it improves on the baseline by nearly 80%. Although the fit between the observed and predicted deletion rates is not perfect, we take this substantial improvement as evidence that our model performs better than the baseline.

Figure 4: Observed and predicted *t/d*-deletion rates in Columbus English. The broken line indicates the predictions based on the baseline, unscaled HG. The solid line shows the predictions based on the frequency weighted HG with a ρ-value of 5.



4. Geminate devoicing in English borrowings in Japanese

4.1. The data

In this section, we present another case study to show the generality of the model that we developed above. Although Japanese native phonology does not tolerate voiced geminates, these sounds have been introduced into Japanese via borrowings from English. Due to Japanese coda restrictions, closed syllables are frequently borrowed with an epenthetic vowel. Additionally, when the English coda consonant is preceded by a lax vowel, the consonant is often geminated. When the English coda consonant is a voiced obstruent, the combination of these processes results in a voiced geminate. In words that contain another voiced obstruent, the geminate optionally devoices, as in the examples in (19) (all examples from Kawahara 2006:538).

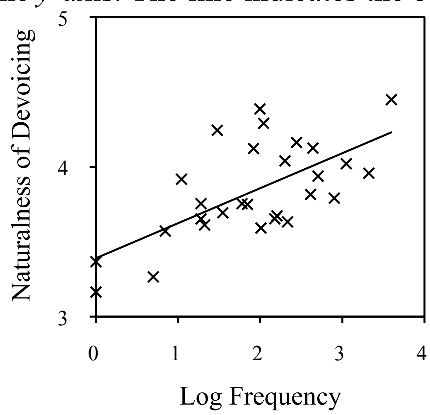
(19)	guddo	~	gutto	'good'
	beddo	~	betto	'bed'
	deibiddo	~	deibitto	'David'
	doggu	~	dokku	'dog'
	baggu	~	bakku	'bag'
	doraggu	~	dorakku	'drug'
	biggu	~	bikku	'big'

This optional devoicing in loanwords has received a lot of attention in recent years so that the factors that condition its application are now well understood. We refer the reader to the literature for a discussion of these factors (Crawford 2009; Kaneko & Iverson 2009; Kawahara 2005, 2006, 2008, 2010, to appear; Nishimura 2003, 2006; Tanaka 2009). Our focus here will be on how this process is influenced by usage frequency. In two recent studies, Kawahara has found a strong positive correlation between geminate devoicing and word frequency (Kawahara 2010, to appear). We will develop an account for the results of Kawahara (2010) here. We summarize the most important aspects of his results below, and refer the reader to the original paper for more details on the design of the experiment.

Kawahara presented 52 native Japanese speakers with 28 loan words like those in (19) with the task of rating the naturalness of a pronunciation in which the voiced geminate has been devoiced. Participants indicated their responses on a 5-point scale, with [5] corresponding to "very natural", and [1] to "very unnatural". The raw usage frequency of each loan word token was taken from the Amano and Kondo Japanese lexical corpus (Amano & Kondo 2000), and log transformed. Figure 5 plots the average naturalness rating that each token received against its log-transformed frequency. Performing a linear regression on these data confirms that log frequency and naturalness are positively correlated ($r^2 = .43$, $p < .001$).

The best way to collect data on devoicing rates in actual speech production would be to investigate the prevalence of devoicing in a large, phonetically transcribed, corpus of spoken Japanese – similar to how we investigated the prevalence of *t/d*-deletion in the *Buckeye Corpus* above. Unfortunately, no such corpus that is large enough to contain enough examples of loanwords exists for Japanese. The second best option would be to conduct a production experiment, designed to collect data on loanwords. Participants in such experiments usually use a rather formal speech style in which optional lenition processes, such as geminate devoicing, is often inhibited. We therefore work under the assumption that naturalness ratings such as those in Kawahara (2010) originate in the same grammar that governs speech production, and that these naturalness ratings therefore also reflect the frequency with which devoicing will apply to the loanwords in actual speech. Even if this is accepted, it is still necessary to convert the 5-point naturalness scale to percent devoicing in some manner. Little is known about how naturalness ratings are related to production patterns (though see Kempen & Harbusch 2008 for some ideas involving syntactic data), and we therefore explored several different options for transforming the naturalness ratings of Kawahara (2010) to percent devoicing. In all of the transformations that we explored, the positive correlation between frequency and rate of devoicing was preserved. We report here on three of the transformations that we considered.

Figure 5: The relation between frequency and devoicing in Kawahara (2010). The *x*-axis represents log-transformed frequencies from Amano and Kondo (2000). The naturalness rating of devoicing is plotted on the *y*-axis. The line indicates the best-fit linear regression line.



The first transformation is a straightforward linear transformation. The assumption is that a rating of [5] corresponds to a token that is always produced with devoicing, a rating of [4] to a token that is produced with devoicing 4/5 of the time (i.e. with 80% devoicing), etc. The second transformation is a variant of the first, the only difference being that the ratings are first replaced by their exponentials. Following a suggestion by Kempen and Harbusch (2008), we also include a transformation where the naturalness ratings and the production patterns are related via a sigmoid function. The formulas used in these three transformations are given in (20). Figure 6 plots the deletion rates under these different transformations against the log frequency of the tokens. The correlation between frequency and devoicing is preserved under these transformations – under all three transformations the correlation is positive and significant (linear: $r^2 = .43$, $p < .001$; exponential: $r^2 = .34$, $p < .002$; sigmoid: $r^2 = .41$, $p < .001$).

(20) Let r be the average naturalness rating that some token t received, and *devoice*(t) the rate of devoicing in token t. Let $norm^r$ be the standardized value of r.

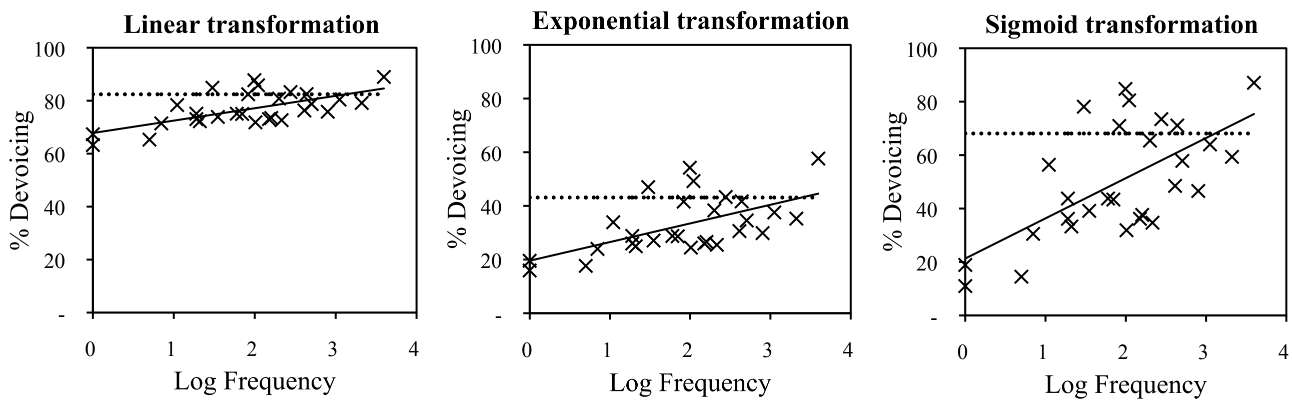
a. Linear transformation: $$devoice(t) = \left(\frac{r}{5}\right)(100)$$

b. Exponential transformation: $$devoice(t) = \left(\frac{e^r}{e^5}\right)(100)$$

c. Sigmoid transformation: $$devoice(t) = \left(\frac{1}{1 + e^{-norm^r}}\right)(100)$$

To determine the overall devoicing rate under each of these transformations, we created a corpus for each transformation, assuming that each loanword appears in the corpus with its frequency in Amano and Kondo (2000). The loanword /deibiddo/ ‘David’, for instance, has a frequency of 83 in Amano and Kondo, and /deibiddo/ was hence represented 83 times in our corpora. Each token was represented with devoicing according to the transformations given in (20). Devoicing in /deibiddo/ received an average rating of 4.12. Performing the linear transformation on this score results in a devoicing rate of 82.4%, and this percentage of the 83 occurrences of /deibiddo/ in the linear corpus was hence represented with devoicing (i.e. 68 tokens with and 15 without devoicing). The same was done for all loanwords in all three corpora. The overall devoicing rate in each of the three corpora was then calculated. This overall rate is shown with broken lines in Figure 6 (linear: 82.4%; exponential: 43.1%; sigmoid: 68.1%). As with the overall rate of *t/d*-deletion in the *Buckeye Corpus* (see Figure 1), the overall rate of devoicing under each transformation is closer to the rate observed for the more frequent words.

Figure 6: The relationship between frequency and percent geminate devoicing under the three different transformations. The solid lines represent the result of linear regressions. The broken lines represent the overall devoicing rate under each transformation.



We will assume these three corpora as data in the rest of this section. As with *t/d*-deletion, we first develop purely grammatical models based on the overall devoicing rate in each corpus, following the tradition in the current literature on phonological variation. We then augment these models with weight scaling according to the method described above in sections 2.2 and 2.3.

4.2. A grammatical account

We rely on the three constraints in (21). As with *t/d*-deletion, we used the noisy HG learning algorithm in *Praat* to learn the weights associated with these constraints for the three corpora. Each input file contained 100 tokens, with the proportion of tokens represented with devoicing determined by the overall rate of devoicing in each corpus.[12] These learning files were submitted to *Praat*'s learning algorithm, using all of the default settings in *Praat*. Once the grammars have been learned, the "to output distribution" function in *Praat* was used to determine the predicted rate of devoicing for each of the three learned grammars. The constraint weights that were learned for each corpus are given in (22), and the predicted rate of devoicing in (23). Unsurprisingly, the match between the observed overall devoicing rates and the devoicing predicted by the grammars is very good. However, as before, these grammars produce devoicing at the overall rates, and also treat all words of all frequencies the same. As with *t/d*-deletion, the overall corpus rates are heavily influenced by frequent words so that the devoicing rates

[12] The input files used for each corpus are available with this submission.

predicted by the grammars are relatively accurate for frequent words, but differ quite drastically from the actual devoicing rates for less frequent words. This is shown in Figure 7.

(21) *GEMINATE — Assign one violation mark for every consonant linked to two timing slots.

*VOICEDOBS — Assign one violation mark for every voiced obstruent.

IDENT[voice] — Assign one violation mark for every output segment that has a different specification for the feature [voice] than its input correspondent.

(22)

	Linear	*Exponential*	*Sigmoid*
*GEMINATE	100.0	100.0	100.0
*VOICEDOBS	101.3	99.7	100.7
IDENT[voice]	98.7	100.3	99.3

(23)

	Observed devoicing (%)	*Expected devoicing (%)*
Linear	82.4	82.2
Exponential	43.1	42.4
Sigmoid	68.1	68.6

4.3. Incorporating the frequency bias through weight scaling

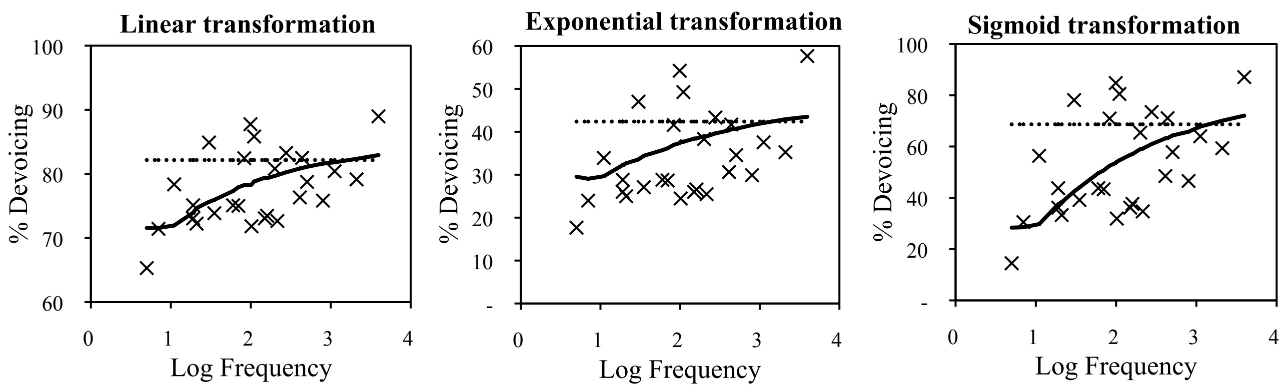
In order to remedy the shortcomings of the purely grammatical accounts above, we incorporate the contribution of usage frequency into these models, in the same way as we did for *t/d*-deletion in section 3.3. What is required is to scale the weight of the faithfulness constraint (IDENT[voice]) up for infrequent words so that they are more likely to be treated faithfully, and conversely to scale the weight of IDENT[voice] down for frequent words. First, we determined the reference point between frequent and infrequent words according to the method described in (9). In total, each corpus contains 11,000 tokens. The two most frequent words account for over half of the 11,000 tokens (/baguddado/ 'Baghdad', frequency: 3,951;[13] /baggu/ 'bag', frequency: 2,103). The reference point is hence halfway between the log frequency of /baggu/ (3.32) and the log frequency of the next most frequent word, /bajji/ 'budge' (3.05), or 3.19. With this reference value in hand, the *beta* distribution associated with each word can now be determined. For all words, the value of α is the reference log frequency of 3.19, and the value of β is the log frequency of the specific word. The value of the range parameter ρ is set to maximize the fit of the model predictions with the data being modeled, exactly as it was done for *t/d*-deletion above in section 3.3. Since we are modeling three different corpora, based on three different grammars learned, the value ρ was determined separately for each of the corpora. For the corpora based on the linear and exponential transformations, the value of ρ resulting in the best fit to the data was found to be 1. For the sigmoid corpus, it was found to be 3.

---

[13] The high frequency of /baguddado/ in Amano and Kondo is a result of their frequency counts being taken from a corpus of newspapers from the time after the American invasion of Iraq. Although it is not clear that /baguddado/ will still have such a high frequency for the average Japanese speaker, we opted not to adjust its frequency for the purposes of this paper. The fact that /baguddado/ pronounced with devoicing, i.e. as [baguttado], received a high naturalness rating in Kawahara (2010) suggests that this choice might be correct.

Once the value of ρ for a corpus has been determined, the scaling factor associated with each word in that corpus can be determined. The weight of the faithfulness constraint can then be scaled according to this scaling factor for each word, and the predicted rate of devoicing can be determined for individual words using the "to output distribution" function in *Praat*. Figure 7 shows how the baseline, unscaled HG models compare with the scaled model for each of the three corpora. The broken line in each graph plots the predictions of the baseline model, and the solid line the predictions of the scaled model. This figure clearly shows that the scaled models fit the data better. This is confirmed by (24), where the mean square error of the baseline and the scaled models relative to the actual devoicing rate is given for each corpus. For all three corpora, the scaled grammar shows a significant improvement over the unscaled grammar.

Figure 7: Observed and predicted devoicing rates for different transformations. The broken lines indicate the predictions based on the basic, unscaled HG. The solid lines show the predictions based on the frequency weighted HG with the best ρ-value for each transformation.



(24) Mean square errors and percent improvement relative to the baseline, unscaled grammars

	Linear		*Exponential*		*Sigmoid*	
	Base	*Scaled (ρ = 1)*	*Base*	*Scaled (ρ = 1)*	*Base*	*Scaled (ρ = 3)*
MSE	52.7	24.5	169.3	88.0	619.5	287.2
Improvement	–	53.5%	–	48.0%	–	42.1%

In addition to the three transformations from naturalness ratings to percent devoicing that we discussed here, we also explored several others. In every transformation where the positive correlation between frequency and devoicing was preserved, the improvement of the weight scaled HG model was comparable to the improvement rates reported in (24). Although the data that we modeled here were not based on actual production, we take this as evidence that a weight scaled model will do significantly better than an unscaled model on actual production data, as long as such data show a comparable positive correlation between frequency and devoicing.

5. Discussion

5.1. Biases other than frequency

We have focused on usage frequency as a non-grammatical factor that influences variation. However, as we noted in section 1.2, frequency is only one of many such factors. In principle, any of these other non-grammatical factors could be incorporated into the model that we propose above in the same way as the influence of frequency has been incorporated; that is, via scaling the weight of faithfulness constraints. However, even if this is accepted, there are many different ways in which this could be implemented. Choosing between different options requires more detailed information on how different non-grammatical factors interact with each other and with the grammar than is currently available. We therefore opt not to take a strong stance, but we rather point out some possible implementations, and the implications of each.

One option is that there is only one scaling factor that jointly captures the contribution of all non-grammatical factors. This could be implemented by having different factors add to the values of the shape parameters, α and β, of the *beta* distribution. In (25), we show how this could be done. Recall from section 2.3 that a *beta* distribution with $\alpha > \beta$ has a positive mode. Any non-grammatical factor that results in an inhibiting bias (that decreases the likelihood of a simplification process applying) could then increase the value of α to increase the likelihood that $\alpha > \beta$, and that the scaling factor will be positive. On the other hand, a bias that promotes the likelihood of simplification could increase the value of β, which would lead to increasing the likelihood of $\alpha < \beta$, and hence of a negative scaling factor.

(25)

Parameter	*Value*
α	log(mid point frequency) + inhibiting bias
β	log(word frequency) + promoting bias

Such an approach has two implications. First, it implies that the non-grammatical factors are not independent from each other. The effect of frequency, for instance, would not necessarily be the same in a formal as opposed to an informal speech situation. Secondly, this approach implies that the influence of non-grammatical factors on variation is limited. The *beta* distribution has a finite range, with the range parameter ρ setting the upper and lower bounds of the range. If there is only one scaling factor, the weight of faithfulness constraints can be scaled at most by 2ρ units. An approach with only one scaling factor therefore affords grammar qualitatively more importance than the non-grammatical factors.

The other logical possibility is that there are several scaling factors, one for each of the non-grammatical factors that impact variation. There would then be a scaling factor for frequency, one for speech register, one for speech rate, etc. The effective weight of faithfulness constraints would be determined by adding the different scaling factors to the weight of the faithfulness constraints. Such an approach rests on the assumption that the different non-grammatical factors are independent from each other. Since there are separate scaling factors for frequency and register, frequency should have the same effect in a formal and an informal speech situation. Secondly, such an approach does not place an upper limit on the influence of non-grammatical factors on variation. Each non-grammatical factor can scale the weight of faithfulness constraints by 2ρ units. But there is no principled limit on the number of non-grammatical factors, and hence no limit on how many units faithfulness weights can be scaled by non-grammatical factors.

Although more data would be necessary to decide definitively between these options, our current preference is for the first option laid out above. Both the variationist literature of the past four decades and the more generative literature on phonological variation have consistently found that grammatical factors strongly determine the patterns of variation. We therefore opt for an approach that limits the contribution of non-grammatical factors.

5.2. Grammar dominance

Although the model that we propose in this paper allows non-grammatical factors to influence phonological variation, it is a grammar dominant model. Grammar sets the limits of what patterns of variation are possible, and all that the non-grammatical factors can do is to determine how variation is realized within these limits. This grammar dominance follows from the fact that our model is developed within HG, which is by definition a grammatical model. The dominance of grammar realizes itself in both universal terms and in the grammars of individual languages.

First consider the universal aspects of grammar dominance. In HG (as in OT), Universal Grammar is represented in the constraint set. Classic OT (Prince & Smolensky 1993, 2004) assumes that the constraint set is universal, so that the grammar of every language contains exactly the same constraints. From this follows that there are certain logically possible grammatical constraints that just do not exist, and if some constraint does not exist then some logically possible grammatical patterns are impossible. For example, in our analysis of *t/d*-deletion, we proposed positional MAX constraints for pre-vocalic and pre-pausal position, but argued that no such positional constraint exists for pre-consonantal position. If this is a true observation about the constraint set, it follows that deletion in pre-consonantal context will always violate only a subset (MAX) of the faithfulness constraints violated by deletion in pre-vocalic (MAX, MAX-PRE-V) or pre-pausal (MAX, MAX-PRE-PAUSE) position. In (26), we show the consequences that this has for the H-score of deletion candidates in the different contexts. The H-score of deletion in pre-consonantal position will always be higher than that of deletion in the other two contexts. This effect cannot be overridden by weight scaling in our model, since we assume that all faithfulness constraints are scaled by the same factor. In any language, under any set of non-grammatical factors incorporated into the model via scaling factors, it follows that deletion will always be most likely in pre-consonantal position. All that the non-grammatical factors can do is to increase the likelihood of deletion in all three contexts, but they will do so to the same extent in all three contexts.[14]

(26)

	MAX	MAX-PRE-V	MAX-PRE-PAUSE	H
/wɛst bæŋk/ → [wɛs_ bæŋk]	-1			-*w*(MAX)
/wɛst ɛnd/ → [wɛs_ ɛnd]	-1	-1		-*w*(MAX) - *w*(MAX-PRE-V)
/wɛst/ → [wɛs]	-1		-1	-*w*(MAX) - *w*(MAX-PRE-PAUSE)

A similar point can be made with regard to geminate devoicing in Japanese. In our analysis, we assumed a markedness constraint that penalizes voiced obstruents, but no constraint that penalizes voiceless obstruents. If no constraint against voiceless obstruents exists, it follows that a grammar that will result in context free voicing of obstruents (whether as a categorical or variable process) is impossible. It does not matter how frequent a word is or in what non-

[14] Since a process cannot apply at a rate of higher than 100%, this statement should be qualified somewhat. Imagine a grammar where pre-consonantal context has a base deletion rate of 80% and pre-pausal context of 50%. Deletion in pre-consonantal position can be increased by at most 20% by the contribution of scaling factors. The same holds for scaling factors that reduce the application of a simplification process and the floor of application, 0%.

grammatical context it is uttered, since this process is ruled out by the grammar, it is predicted never to be observed.

Grammar also takes precedence over usage frequency and other non-grammatical factors at the level of individual languages. In the grammar developed for Columbus English above, the weight of MAX-PRE-V (-1.51) is lower than that of MAX-PRE-PAUSE (0.96), corresponding to the fact that this dialect of English shows more deletion in pre-vocalic than pre-pausal position. Since the weight of all faithfulness constraints are scaled by the same amount, the relative difference in the effective weights of MAX-PRE-V and MAX-PRE-PAUSE will be preserved under all scaling conditions. No matter how frequent a specific word is or under what non-grammatical conditions it is uttered, on average a pre-vocalic deletion candidate will have a higher H-score than a pre-pausal deletion candidate. The grammar of Columbus English stipulates that deletion is more likely in pre-vocalic context, and non-grammatical factors cannot override this.

This dominance of grammar depends on the assumption that at a given instance of using the grammar (evaluation of a specific word, at a specific instance, under a specific set of non-grammatical circumstances) the weights of *all* faithfulness constraints are scaled by the same amount. If weight scaling could variably affect different faithfulness constraints, the dominance of grammar could be lost. In this regard, our proposal diverges from the related proposal made by Boersma and Hayes in Appendix C of their 2001 paper. Their model is developed in stochastic OT, and they therefore assume constraint ranking rather than weighting. They propose that the ranking values of some constraints can be changed in different speech situations. But crucially, they propose that some constraints can be ranked higher, others lower, and that constraint rankings do not have to be changed by the same amount. More data on how grammar and non-grammatical factors interact would be necessary to decide definitively between these two options. However, in the absence of sufficient data, we opt for the more restrictive model.

The dominance of grammar is also not a property of all other models of phonological variation. In some implementations of usage-based models (Bybee 2001, 2006; 2007; etc.), or exemplar models (Gahl & Yu 2006 and papers therein; Pierrehumbert 2001; etc.), no formal distinction is made between grammatical and non-grammatical factors. In fact, in describing usage-based grammar in her presidential address to the *Linguistic Society of America*, Bybee first defines the usage-based conceptualization of grammar as "the cognitive organization of one's experience with language" (Bybee 2006:711). Later on the same page she describes how this organization is done as follows: "… the general cognitive capabilities of the human brain, which allow it to categorize and sort for identity, similarity, and difference, go to work on the language events a person encounters, categorizing and entering in memory these experiences." Grammar is the result of cognitive organization achieved with general cognitive abilities, not with grammar or language specific abilities. Exactly the same cognitive abilities that organize our experience with social interactions and with our physical environment organize our experience with language. No formal distinction is made between how language and other aspects of our experience are processed or stored in the mind. If a child acquiring a language were to be exposed to a set of experiences where deletion happens to be observed more often in pre-vocalic than pre-consonantal context, the general abilities of the mind to classify would notice this pattern, and codify this as the grammar. This is fundamentally different from the type of approach that we advocate above. Under our approach, there are language specific cognitive capacities (Universal Grammar represented in the constraint set, as well as the principles for how constraints interact via their weights). Language is processed according to these principles and not with our general cognitive capabilities. This places a limit on the types of grammars that could be learned. As we showed above, the assumptions about Universal Grammar under which we operate implies that no grammar that produces more deletion in pre-vocalic than pre-consonantal context is possible.

Also on this question, more research is necessary to determine to what extent certain types of grammars are truly impossible. A long tradition of typological research has established strong universal patterns across languages, a result that could be interpreted as favoring a system that includes a strong Universal Grammar. Recent research in artificial grammar learning has also shown that linguistic patterns that counter such universal trends are either unlearnable or at least not easily learnable (Carpenter 2006; Coetzee 2009b; Pater & Tessier 2006; Wilson 2006). On the other hand, there are also unambiguous examples of languages with grammars that counter universal trends (Coetzee & Pretorius 2010; Hyman 2001), showing that it should be possible for language learners to acquire grammars that do not fit neatly into the constrains of Universal Grammar. Along similar lines, Bybee (2002:275) also shows that in one dialect of English some words, under some circumstances, show more word-final *t/d*-deletion in pre-vocalic than pre-consonantal context. With conflicting data from the current literature it is impossible to choose definitively between a model with a strong grammar, dominating other factors, and a model in which grammar is afforded no special place. However, given that the evidence for strong universal tendencies is currently more copious than evidence for linguistic systems that counter these tendencies, we opt for the more restrictive model where Universal Grammar places limits on possible languages.

5.3. Final summary

In this paper, we developed a model of phonological variation that incorporates influences from both grammatical and non-grammatical factors. Our model retains some of the core characteristics of a classical generative grammar, while also embracing insights from usage-based and exemplar models of grammar. In the phonological literature, the generative approach and the usage-based/exemplar approaches have often been presented as opposites and as incompatible with each other. We believe this to be a false dichotomy. Not only is it possible to integrate these approaches seamlessly, but such an integration also enables phonological theory to account better for many phenomena than what either of the two approaches could do in isolation. If such an integration is indeed the correct route to go, then future research will have to focus on two issues. First, the proper way to integrate the contributions from the two types of models needs to be determined. This paper contains one proposal, and the success of this proposal leads us to believe that it has merit. But other ways of integration are possible, and more research is necessary to determine all of the viable options, and to evaluate their success. Secondly, more targeted data collection would need to be performed. The data on phonological variation that are currently available are usually not suited to address the questions raised by an integrated model such as that proposed in this paper.

References

Amano, Shigeaki & Tadahisa Kondo. (2000). *NTT Database Series: Lexical Properties of Japanese, 2nd Release*. Tokyo: Sanseido.

Anttila, Arto. (1997). Deriving variation from grammar. In Hinskens, Frans, Roeland van Hout & Leo Wetzels (eds.) *Variation, Change and Phonological Theory*. Amsterdam: John Benjamins, 35-68.

Anttila, Arto. (2002a). Morphologically conditioned phonological alternations. *Natural Language and Linguistic Theory* 20, 1-42.

Anttila, Arto. (2002b). Variation and phonological theory. In Chambers, Jack K., Peter Trudgill & Natalie Schilling-Estes (eds.) *Handbook of Language Variation and Change*. Oxford: Blackwell, 206-243.

Anttila, Arto. (2006). Variation and opacity. *Natural Language and Linguistic Theory* 24, 893-944.

Anttila, Arto. (2007). Variation and optionality. In de Lacy, Paul (ed.) *The Cambridge Handbook of Phonology*. Cambridge: Cambridge University Press, 519-536.

Anttila, Arto, Vivienne Fong, Stefan Benus & Jennifer Nycz. (2008). Variation and opacity in Singapore English consonant clusters. *Phonology* 25, 181-216.

Baayen, R. Harald, Richard Piepenbrock & Leon Gulikers. (1995). *The CELEX Lexical Database (CD-ROM)*. Philadelphia: Linguistic Data Consortium.

Baese-Berk, M. & Matt Goldrick. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes* 24, 147-185.

Bayley, Robert. (1995). Consonant cluster reduction in Tejano English. *Language Variation and Change* 6, 303-326.

Bayley, Robert. (2002). The quantitative paradigm. In Chambers, J.K, Peter Trudgill & Natalie Schilling-Estes (eds.) *The Handbook of Language Variation and Change*. Oxford: Blackwell, 117-141.

Bell, Alan, Jason Brenier, Michelle Gregory, Cynthia Girand & Daniel Jurafsky. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60, 92-111.

Boersma, Paul. (1997). How we learn variation, optionality, and probability. *University of Amsterdam Institute of Ponetic Sciences Proceedings* 21, 43-58.

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.] Retreived October 31, 2009, from http://www.praat.org.

Browman, Catherine P. & Louis Goldstein. (1990). Tiers in articulatory phonology, with some implications for casual speech. In Kingston, John & Mary E. Beckman (eds.) *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambdrige: Cambridge University Press, 341-376.

Bybee, Joan L. (2000). The phonology of the lexicon: evidence from lexical diffusion. In Barlow, Michael & Suzanne Kemmer (eds.) *Usage-Based Models of Language*. Stanford: CSLI, 65-85.

Bybee, Joan L. (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.

Bybee, Joan L. (2002). Word frequency and context of use in lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14, 261-290.

Bybee, Joan L. (2006). From usage to grammar: the mind's response to repetition. *Language* 82, 711-733.

Bybee, Joan L. (2007). *Frequency of Use and the Organization of Language*. Oxford: Oxford University Press.

Carpenter, Angela (2006). *Acquisition of a Natural Versus an Unnatural Stress System*. Ph.D. Dissertation, Unversity of Massachusetts.

Celdran, E.M. & X. Villalba. (1995). Locus equations as a metric for place of articulation in automatic speech recognition. *ICPhS XIII: Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm, Sweden. Volume 1*, 30-33.

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*. Ph.D., Dissertation, University of Massachusetts.

Coetzee, Andries W. (2006). Variation as accessing "non-optimal" candidates. *Phonology* 23, 337-385.

Coetzee, Andries W. (2009a). An integrated grammatical/non-grammatical model of phonological variation. In Kang, Young-Se, Jong-Yurl Yoon, Hyunkyung Yo, Sze-Wing Tang, Yong-Soon Kang, Youngjun Jang, Chul Kim, Kyoung-Ae Kim & Hye-Kyung Kang (eds.) *Current Issues in Linguistic Interfaces. Volume 2*. Seoul: Hankookmunhwasa., 267-294.

Coetzee, Andries W. (2009b). Learning lexical indexation. *Phonology* 26, 109-145.

Coetzee, Andries W. (2009c). Phonological variation and lexical frequency. In Schardl, Anisa, Martin Walkow & Muhammad Abdurrahman (eds.) *NELS 38. Volume 1*. Amherst: GLSA, 189-202.

Coetzee, Andries W. (to appear). Variation: where laboratory and theoretical phonology meet. In Cohn, Abigail C., Cécile Fougeron & Marie Huffman (eds.) *The Oxford Handbook of Laboratory Phonology*. Ofxord: Oxford University Press.

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.

Coetzee, Andries W. & Rigardt Pretorius. (2010). Phonetically grounded phonology and sound change: the case of Tswana labial plosives. *Journal of Phonetics* 38, 404-421.

Côté, Marie-Hélène. (2004). Syntagmatic distinctness in consonant deletion. *Phonology* 21, 1-41.

Crawford, Clifford James (2009). *Adaptation and transmission in Japanese loanword phonology*. Ph.D. dissertation, Cornell University.

Diehl, Randy L. & Margaret A. Walsh. (1989). An auditory basis for the stimulus-length effect in the perception of stops and glides. *Journal of the Acoustical Society of America* 85, 2154-2164.

Eek, A. & E. Meister. (1995). The perception of stop consonants: locus equations and spectral integration. *ICPhS XIII: Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm, Sweden. Volume 1*, 18-21.

Fasold, R. (1972). *Tense Marking in Black English*. Arlington: Center for Applied Linguistics.

File-Muriel, Richard J. (2010). Lexical frequency as a scalar variable in explaining variation. *The Canadian Journal of Linguistics* 55, 1-25.

Fowler, Carol A. (1994). Invariants, specifiers, cues: an investigation of locus equations as information for place of articulation. *Perception & Psychophysics* 55, 597-610.

Fruchter, David & Harvey M. Sussman. (1997). The perceptual relevance of locus equations. *Journal of the Acoustical Society of America* 102, 2997-2308.

Gahl, Susanne. (2008). *Time* and *thyme* are not homophones: the effect of lemma frequency on word durations in spontaneous speech. *Language* 84, 474-496.

Gahl, Susanne & Alan Yu. (eds.) (2006). *Linguistic Review. Volume 23(3). Special Issue on Exemplar-Based Models in Linguistics*. Berlin: De Gruyter Mouton.

Goeman, A.C.M. (1999). *T-deletie in Nederlandse Dialecten. Kwantitatiewe Analyse van Structurele, Ruimtelijke en Temporele Variate*. Ph.D. Dissertation, Vrije Universiteit, Amsterdam. The Hague: Holland Academic Graphics.

Goeman, A.C.M. & P.Th. van Reenen. (1985). *Word-Final t-Deletion in Dutch Dialects. The Roles of Conceptual Prominence, Articulatory Complexity, Paradigmatic Properties, Token Frequency and Geographical Distribution*. Amsterdam: Vakgroep Algemene Taalwetenschap Vrije Universiteit.

Goldinger, S.D., Paul A. Luce, David B. Pisoni & J.K. Marcario. (1992). Form-based priming in spoken word recognition: The roles of competition and bias. *Journal of Experimental Psychology: Learning, Memory and Cognition* 18, 1211-1238.

Goldsmith, John. (1993). Harmonic phonology. In Goldsmith, John (ed.) *The Last Phonological Rule: Reflections on Constraints and Derivations*. Chicago: Chicago University Press, 21-60.

Goldsmith, John. (ed.) (1995). *The Handbook of Phonological Theory*. Oxford: Blackwell.

Gupta, Arjun K. & Saralees Nadarajah. (eds.) (2004). *Handbook of the Beta Distribution and Its Applications*. New York: Marcel Dekker.

Guy, Gregory R. (1980). Variation in the group and the individual. In Labov, William (ed.) *Locating Language in Time and Space*. New York: Academic, 1-36.

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

Guy, Gregory R. (to appear). Variability. In van Oostendorp, Marc, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.) *The Blackwell Companion to Phonology*. Oxford: Blackwell.

Guy, Gregory R. & Charles Boberg. (1997). Inherent variability and the obligatory contour principle. *Language Variation and Change* 9, 149-164.

Holmes, Janet. (1995). Glottal stops in New Zealand English: an analysis of variants of the word-final /t/. *Linguistics* 33, 433-463.

Hooper, Joan B. (1976). Word frequency in lexical diffusion and the source of morphological change. In Christie, William M. (ed.) *Current Progress in Historical Linguistics*. Amsterdam: North-Holland Publishing Co., 95-105.

Hyman, Larry. (2001). On the limits of phonetic determinism in phonology: *NC revisited. In Hume, Elizabeth & Keith Johnson (eds.) *The Role of Speech Perception Phenomena in Phonology*. New York: Academic Press, 141-185.

Jesney, Karen. (2007). *The locus of variation in weighted constraint grammars*. Ms. Poster presented at the Workshop on Variation, Gradience and Frequency in Phonology. Stanford University, July 2007. [Downloaded on December 27, 2007 from http://people.umass.edu/kjesney/papers.html.].

Jurafsky, Daniel, Alan Bell, Michelle Gregory & William D. Raymond. (2001). Probabilistic relations between words: evidence from reduction in lexical production. In Bybee, Joan B. & Paul Hopper (eds.) *Frequency and the Emergence of Linguistic Structure*. Amsterdam: Benjamins, 229-254.

Kaisse, Ellen M. & Patricia A. Shaw. (1985). On the theory of Lexical Phonology. *Phonology Yearbook* 2, 1-30.

Kaneko, Emiko & Gregory K. Iverson. (2009). Phonetic and other factors in Japanese on-line adaptation of English final consonants. In Inagaki, S. & M. Hirakawa (eds.) *Studies in Language Sciences 8: Papers from the eighth annual conference of the Japanese Society for Language Science*. Tokyo: Kuroshio Publications.

Kang, Hye-Kyung. (1994). Variation in past-marking and the question of the system in Trinidadian English. In Beals, Katherine (ed.) *CLS 30: Papers from the 30th Regional Meeting of the Chicago Linguistic Society. Volume 2: The Parasession on Variation in Linguistic Theory*. Chicago: Chicago Linguistic Society.

Kawahara, Shigeto. (2005). Voicing and geminacy in Japanese: An acoustic and perceptual study. In Flack, Katherine & Shigeto Kawahara (eds.) *Univeristy of Massachusetts Occasional Papers in Linguistics 31*. Amherst: GLSA, 87-120.

Kawahara, Shigeto. (2006). A faithfulness ranking projected from a perceptibility scale: the case of [+voice] in Japanese. *Language* 82, 536-574.

Kawahara, Shigeto. (2008). Phonetic naturalness and unnaturalness in Japanese loanword phonology. *Journal of East Asian Linguistics* 18, 317–330.

Kawahara, Shigeto. (2010). *Aspects of Japanese Loanword Devoicing*. Ms. Rutgers University. [ROA #1082.].

Kawahara, Shigeto. (to appear). Japanese loanword devoicing revisited: A wellformedness judgment study. *Natural Language and Linguistic Theory*.

Kempen, Gerard & Karin Harbusch. (2008). Comparing linguistic judgments and corpus frequencies as windows on grammatical competence: a study of argument linearization in German clauses. In Steube, Anita (ed.) *The Discourse Potential of Underspecified Structures*. Berlin: Walter de Gruyter, 179–192.

Kewley-Port, Diane. (1983). Time-varying features as correlates of place of articulation in stop consonants. *Journal of the Acoustical Society of America* 73, 322-335.

Kewley-Port, Diane, David Pisoni & Michael Studdert-Kennedy. (1983). Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants. *Journal of the Acoustical Society of America* 73, 1779-1793.

Kiparsky, Paul. (1985). Some consequences of Lexical Phonology. *Phonology Yearbook* 2, 85-138.

Kiparsky, Paul. (1993). *An OT perspective on phonological variation*. Ms. Handout from Rutgers Optimality Workshop 1993, also presented at NWAV 1994, Stanford University. Available at http://www.stanford.edu/~kiparsky/Papers/nwave94.pdf.

Labov, William. (1966). *The Social Stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

Labov, William. (1969). Contraction, deletion, and inherent variability of the English copula. *Language* 45, 715-762.

Labov, William. (1989). The child as linguistic historian. *Language Variation and Change* 1, 85-97.

Labov, William. (2004). Quantitative analysis of linguistic variation. In Ammon, Ulrich, Norbert Dittmar, Klaus J. Mattheier & Peter Trudgill (eds.) *Sociolinguistics: An International Handbook of the Science of Language and Society. Volume 1. 2nd edition.* Berlin: Mouton de Gruyter, 6-21.

Lacoste, Véronique (2008). *Learning the Sounds of Standard Jamaican English: Variationist, Phonological and Pedagogical Perspectives on 7-Year-Old Children's Classroom Speech*. Ph.D. Dissertation, University of Essex.

Lahiri, Aditi, Letitia Gewirth & Sheila E. Blumstein. (1984). A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: evidence from a cross-language study. *Journal of the Acoustical Society of America* 76, 391-404.

Legendre, Géraldine, Yoshiro Miyata & Paul Smolensky. (1990). Can connectionism contribute to syntax? Harmonic Grammar, with an application. In Ziolkowski, M., M. Noske & K. Deaton (eds.) *Proceedings of the 26th Regional Meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society, 237-252.

Lindblom, Bjorn. (1990). Explaining phonetic variation: a sketch of the H & H theory. In Hardcastle, W.J. & A. Marchal (eds.) *Speech Production and Speech Modeling*. Dordrecht: Kluwer Academic Publishers, 403-439.

Luce, Paul A. & David B. Pisoni. (1986). Recognizing spoken words: the neighborhood activation model. *Ear and Hearing* 19, 1-35.

Malécot, André. (1958). The role of releases in the identification of released final stops: a series of tape-cutting experiments. *Language* 34, 274-284.

McCarthy, John J. & Alan Prince. (1995). Faithfulness and reduplicative identity. In Beckman, Jill, Suzanne Urbanczyk & Laura Walsh Dickey (eds.) *University of Massachusetts Occasional Papers in Linguistics 18: Papers in Optimality Theory*. Amherst: GLSA.

McNamara, Timothy P. (2005). *Semantic Priming: Perspectives from Memory and Word Recognition*. New York: Psychology Press, Taylor & Francis.

Mitterer, Holger & Mirjam Ernestus. (2006). Listeners recover /t/'s that speakers reduce: evidence from /t/-lenition in Dutch. *Journal of Phonetics* 34, 73-103.

Nearey, T.M. & S.E. Shammas. (1987). Formant transitions as partly distinctive invariant properties in the identification of voiced stops. *Canadian Acoustics* 15, 17-24.

Newmeyer, Frederick J. (2003). Grammar is grammar and usage is usage. *Language* 79, 682-707.

Nishimura, Kohei (2003). *Lyman's Law in Loanwords*. M.A. thesis, Nagoya University.

Nishimura, Kohei. (2006). Lyman's Law in loanwords. *Phonological Studies [Onin Kenkyuu]* 9, 83-90.

Pater, Joe. (2008). Gradual learning and convergence. *Linguistic Inquiry* 39, 334-345.

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

Pater, Joe & Anne-Michelle Tessier. (2006). L1 phonotactic knowledge and the L2 acquisition of alternations. In Slabakova, Roumyana, Silvina A. Montrul & Philippe Prévost (eds.) *Inquiries in Linguistic Development: Studies in Honor of Lydia White*. Amsterdam: Benjamins, 115-131.

Patrick, Peter L. (1992). Creoles at the intersection of variable processes: *-t*, *d* deletion and past-marking in the Jamaican mesolect. *Language Variation and Change* 3, 171-189.

Patterson, David & Cynthia M. Connine. (2001). Variant frequency in flap production: A corpus analysis of variant frequency in American English flap production. *Phonetica* 58, 254-275.

Phillips, Betty S. (2006). *Word Frequency and Lexical Diffusion*. New York: Palgrave Macmillan.

Pierrehumbert, Janet. (2001). Exemplar dynamics: word frequency, lenition and contrast. In Bybee, Joan B. & Paul Hopper (eds.) *Frequency Effects and the Emergence of Lexical Structure*. Amsterdam: Benjamins, 137-157.

Pitt, Mark A., Laura Dilley, Keith Johnson, S. Kieling, William D. Raymond, Elizabeth Hume & E. Fosler-Lussier. (2007). *Buckeye Corpus of Conversational Speech. (2nd Release). [www.buckeyecorpus.osu.edu]*. Columbus: Department of Psychology, Ohio State University.

Postal, Paul. (1966). Review of "Elements of General Linguistics" by André Martinet. *Foundations of Language* 2, 151-186.

Postal, Paul. (1968). *Apects of Phonological Theory*. New York: Harper & Row.

Prince, Alan & Paul Smolensky. (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*. Ms. New Brunswick, Rutgers University. [Revised version appeared as Prince & Smolensky 2004.].

Prince, Alan & Paul Smolensky. (2004). *Optimality Theory: Constraint Interaction in Generative Grammar*. Oxford: Blackwell. [Revised version of Prince & Smolensky 1993.].

Raymond, William D., Robin Dautricourt & Elizabeth Hume. (2006). Word-medial /t,d/ deletion in spontaneous speech: modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change* 18, 55-97.

Reynolds, Bill (1994). *Variation and Phonological Theory*. Ph.D. Dissertation, University of Pennsylvania.

Santa Ana, Otto (1991). *Phonetic Simplification Processes in the English of the Barrio: A Cross-Generational Sociolinguistic Study of the Chicanos of Los Angeles*. Ph.D. dissertation, University of Pennsylvania.

Scarborough, Rebecca (2004). *Coarticulation and the Structure of the Lexicon*. Ph.D. Dissertation, UCLA.

Scarborough, Rebecca. (to appear). Lexical and contextual predictability: confluent effects on the production of vowels. In Fougeron, Cécile, B. Kühnert, M. D'Imperio & N. Vallée (eds.) *Papers in Laboratory Phonology X: Variation, Phonetic Detail and Phonological Modeling*. Berlin: Mouton de Gruyter.

Schouten, M.E.H. (1982). T-deletie in de stad Utrecht: schoolkinderen en grootouders. *Forum der Letteren* 23, 282-291.

Schouten, M.E.H. (1984). T-deletie in het zuiden van die provincie Utrecht. *Taal en Tongval* 36, 162-173.

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*. Cambrdige, MA: MIT Press.

Steriade, Donca. (1999). Phonetics in phonology: The case of laryngeal neutralization. In Gordon, Matthew K. (ed.) *UCLA Working Papers in Linguistics 2. (Papers in Phonology 3)*. Los Angeles: Department of Linguistics, UCLA, 25–146.

Steriade, Donca. (2001). Directional asymmetries in place assimilation. In Hume, Elizabeth & Keith Johnson (eds.) *The Role of Speech Perception in Phonology*. San Diego: Academic Press, 219-250.

Steriade, Donca. (2008). The phonology of perceptibility effects: the P-map and its consequences for constraint organization. In Hanson, Kristin & Sharon Inkelas (eds.) *The Nature of the Word: Studies in Honor of Paul Kiparsky*. Cambridge: MIT Press, 151-180.

Stevens, Kenneth N. & Sheila E. Blumstein. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America* 64, 1358-1368.

Stevens, Kenneth N. & Samuel J. Keyser. (1989). Primary features and their enhancement in consonants. *Language* 65, 81-106.

Sussman, Harvey M., Helen A. McCaffrey & Sandar A Matthews. (1991). An investigation of locus equations as a source of relational invariance for stop place categorization. *Journal of the Acoustical Society of America* 90, 936-946.

Tanaka, Shin-Ichi. (2009). The eurhythmics of segmental melody. *Journal of the Phonetic Society of Japan* 13, 44-52.

van Oostendorp, Marc. (1997). Style levels in conflict resolution. In Hinskens, Frans, Roeland van Hout & Leo Wetzels (eds.) *Variation, Change and Phonological Theory*. Amsterdam: John Benjamins, 207-229.

Versace, R. & B. Nevers. (2003). Word frequency effect on repetition priming as a function of prime duration and delay between the prime and the target. *British Journal of Psychology* 94, 389-408.

Vitevitch, Michael S. & Paul A. Luce. (1998). When words compete: levels of processing in spoken word recognition. *Psychological Science* 9, 325-329.

Vitevitch, Michael S. & Paul A. Luce. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language* 40, 734-408.

Walsh, Margaret A. & Randy L. Diehl. (1991). Formant transition duration and amplitude rise time as cues to the stop/glide distinction. *The Quaterly Journal of Experimental Psychology* 43A, 603-620.

Wilson, Colin. (2006). Learning phonology with a substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30, 945-982.