A shorter version of this paper will appear in the *Proceedings of NELS 38*. The only difference between this longer version and the shorter *NELS* version, is that the *NELS* version does not contain the two appendices which are included in this version of the paper.

Phonological Variation and Lexical Frequency

Andries W. Coetzee

University of Michigan

February 2008

Abstract. Lexical usage frequency is known to influence the application rate of some variable processes. Specifically, variable lenition processes typically affect frequent lexical items more often than infrequent lexical items. For instance, variable *t/d*-deletion in English is more likely to apply to a frequent word (*just*) than to an infrequent word (*bust*). Existing grammatical models of phonological variation do well at accounting for the influence of grammatical factors on variation, but these models cannot account for the contribution of non-grammatical factors such as lexical frequency. In this paper, I propose a model of phonological variation that can simultaneously account for the influence of both these factors on variation. I assume an Optimality Theoretic grammar with lexically indexed faithfulness constraints. Variation arises as a result of variable lexical indexation – a single lexical item can be assigned to different lexical classes on different evaluation occasions, and will hence not always be evaluated by the same faithfulness constraints. Each lexical item is associated with a probabilistic distribution function that determines the likelihood of it being assigned to each of the lexical classes. The shape of a lexical item's distribution function is determined by its usage frequency, so that frequency influences the likelihood of the lexical item being assigned to specific lexical classes, and hence the likelihood of it being evaluated by specific faithfulness constraints. I apply the proposed model to variable *t/d*-deletion in English, and show that it succeeds in accounting for the way in which usage frequency influences this process.

In two appendices, I show how the model can be implemented by interpreting the lexical distribution functions as instantiations of the *beta* distribution (Evans *et al.* 2000). This implementation of the model is then used to determine the expected *t/d*-deletion rates in a corpus of American English. I also use this implementation to model the acquisition path of a variable process, showing that the model predicts more variation during earlier acquisitional stages.

Phonological Variation and Lexical Frequency[*]

Andries W. Coetzee

University of Michigan

1. The Lexicon in Variation

During the first few decades after the publication of SPE (Chomsky & Halle 1968), phonological variation was regarded as a phenomenon that is limited to the late stages of phonological derivation.[1] In Lexical Phonology (Kiparsky 1982), for instance, it was assumed that lexical rules apply obligatorily while "postlexical rules can be optional and subject to variation due to rate of speech" (Kaisse & Shaw 1985:6).[2] Kiparsky (1985:86) uses English nasal place assimilation to illustrate this: Intra-morphemically, coda nasals must agree in place with a following consonant, but across word boundaries assimilation is optional.

(1) **English nasal place assimilation**

a. Intra-morphemically = Lexical = Obligatory

e[nt]*er*, **e*[mt]*er*, **e*[ŋt]*er*

a[mb]*er*, **a*[nb]*er*, **a*[ŋb]*er*

pra[ŋk], **pra*[nk], **pra*[mk]

b. Across word-boundaries = Postlexical = Optional

gree[n b]*ox* ~ *gree*[m b]*ox*

i[n b]*ed* ~ *i*[m b]*ed*

gree[n k]*ard* ~ *gree*[ŋ k]*ard*

---

[*] I am greatly indebted to Joe Pater for many insightful discussions of the ideas that went into this paper. I also acknowledge the input of Pam Beddor, John McCarthy, John Kingston, Abby Cohn, Ingvar Lofstedt, and the audiences at MCWOP 11, NELS 38, Stanford Workshop on Gradience and Variation in Phonology, and the Michigan State University.

[1] This section is based largely on Coetzee and Pater (2008).

[2] See also Donegan and Stampe (1979:145) for the related claim that Natural Phonology's processes, but not its rules, can be optional.

Over the years, more and more examples of variation were discovered that crucially depend on morphology, which lead to the realization that variation cannot be limited to late stage phonetic implementation rules. Since I will focus on *t/d*-deletion in the rest of this paper, I discuss this process here as an example. However, there are other equally influential examples, such as variable reduplication in Ilokano (Hayes & Abad 1989) and variable realization of the genitive plural in Finnish (Anttila 1997, 2002).

English variably deletes *t/d* from word final consonant clusters so that a word like *west* can be pronounced as either [wɛst] or [wɛs]. This is probably the most widely studied variable phonological process with the consequence that we understand the factors that influence the application of this process rather well. Labov (1989) and Coetzee (2004: Chapter 5) review the factors that are known to influence this process. One of the factors that influence the likelihood of deletion is the morphological status of the word-final *t/d*. Guy (1991a) points out that deletion is most likely to apply when the *t/d* is part of a monomorpheme (*mist*), less likely when it is the past tense suffix in a semi-weak verb (*kept*), and least likely when it is the sole marker of the past tense in a regular past tense form (*missed*). The table in (2) lists the deletion rates in these three contexts for a few English dialects, but see Guy (1994) and Labov (2004:15-16) for further evidence of the robustness of this generalization.

(2) **Deletion rate of *t/d* in different dialects of English**

	Regular past (*missed*)	*Semi-weak past* (*kept*)	*Monomorpheme* (*mist*)
Philadelphia English (Guy 1991b)	17%	34%	38%
Chicano English (Santa Ana 1992)	26%	41%	58%
Tejano English (Bayley 1997)	24%	34%	56%

With the discovery of more examples like these came the realization that variation had to be dealt with as an integral part of phonological grammar. Not all variation could be relegated to phonetic implementation. This lead to the development of several models of phonological grammar that formally incorporated variation into the grammar (Anttila 1997 *et seq.*; Boersma 1998; Boersma & Hayes 2001; Coetzee 2006; Labov 1969; Reynolds 1994; etc.). Generally speaking, these models were fairly successful at accounting for the influence that grammar has on the application of variable processes.

However, there is another factor that is known to influence variation, and for which the existing models cannot adequately account: the lexicon. Specifically, variable lenition processes such as *t/d*-deletion typically apply more frequently to words with higher usage frequency than to words with lower usage frequency (Bybee 2002; Hooper 1976; etc.). Unfortunately, we do not have detailed information on how usage frequency impacts *t/d*-deletion. In the literature on *t/d*-deletion there is some informal acknowledgement that usage frequency is relevant to the application of this process.

Many studies exclude forms like *and*, *n't*, *went* and *just* from their data since these forms typically show anomalously high rates of deletion. Bybee (2000:70) and Patrick (1992:172) both point out that these are words with very high usage frequency. In order to illustrate the influence of usage frequency on *t/d*-deletion, Bybee (2000) reanalyzed the Chicano English corpus originally collected by Santa Ana (1991). She selected from the corpus 2049 tokens of words that end on /-Ct, -Cd/, and divided the tokens into two groups based on their text frequency in Francis and Kučera (1982). The "high frequency" group all had occurrences of 35 or above per million, and the "low frequency" group less than this. She found a significant correlation between frequency and likelihood of deletion, as shown in (3).

(3) **Rate of *t/d*-deletion in Chicano English**

	Deletion	*Retention*	*% Deletion*
High frequency	898	752	54.4%
Low frequency	137	262	34.4%

The usage frequency of a lexical item cannot be deduced from its phonological properties. It is an idiosyncratic property the lexical item, and therefore needs to be stored with the lexical entry of each individual lexical item. Existing models of phonological variation are all strictly grammatical and can interact with lexical items only in terms of their grammatical properties (their phonological and morphological properties). These models lack the ability to incorporate non-grammatical factors, such as usage frequency, formally into the derivation of an output.

In the rest of this paper, I develop a model of phonological variation that simultaneously accounts for the influence of phonological grammar and lexically specific information such as usage frequency. In this model, I assume an Optimality Theoretic type grammar with faithfulness constraints indexed to lexical classes (Itô & Mester 1999; Pater 2000). I diverge from the literature on lexically indexed constraints by proposing that the lexical class affiliation of specific lexical items is not fixed. Lexical items can vary in their class affiliation from one evaluation occasion to the next, potentially resulting in variable output selection. I also propose that each lexical item is associated with a probabilistic distribution function that determines the likelihood with which it is assigned to each of the available lexical classes. Usage frequency is enabled to influence variation via the lexical distribution functions – frequency is the main determinant of the shape of the distribution functions. Section §2 below is dedicated to describing the data that I will analyze, as well as developing an account for the way in which grammatical factors impact the variation. In section §3, I discuss how usage frequency influences the variation via the lexical distribution functions.

2. Accounting for the role of grammar

There are many grammatical factors that impact the likelihood of *t/d*-deletion, including the morphological status of the *t/d* (see above), the identity of the consonant preceding the *t/d*, what follows on the *t/d*, whether *t/d* appears in a stressed or unstressed syllable, etc. I will focus here only on the influence of the following context. In §2.1, I describe the data that I analyze and the constraints that I use. Section §2.2 presents the analysis.

2.1 Influence of the following context on *t/d*-deletion

Labov (1989) reviews the literature on *t/d*-deletion, and points out that in every dialect of English in which this process has been studied, deletion is most likely to apply in pre-consonantal position (*west bank*).[3] Pre-vocalic (*west end*) and phrase-final (*west##*) positions usually show lower deletion rates, with dialects varying in which of these positions most resists deletion. In (4), I give one example of each of the two kinds of dialects. See Coetzee and Pater (2008) and Coetzee (2004: Chapter 5) for more examples.

(4) ***t/d*-deletion rate in different contexts**

	Pre-C	*Pre-V*	*Pre-##*
Chicano (Santa Ana, 1991)	62%	45%	37%
Tejano (Bayley, 1995)	62%	25%	46%

I propose the three markedness constraints in (5) to account for this pattern of deletion. These constraints are inspired by Steriade's (2001, to appear) "licensing by cue" constraints. The perceptual cues for a consonant are mostly realized in the consonantal release and the formant transitions from the consonant into a following vowel. Since neither of these can be realized before a consonant, *Ct#C universally ranks highest. The other two constraints can vary in their ranking depending on whether or not phrase final stops are released and on how likely formant transitions are to be realized across word boundaries. See Coetzee (2004:221-228) for a detailed motivation of these constraints. The relevant faithfulness constraint is MAX.

(5) *Ct#C No word-final [-Ct]/[-Cd] followed by a [C-] initial word.
*Ct#V No word-final [-Ct]/[-Cd] followed by a [V-] initial word.
*Ct## No word-final [-Ct]/[-Cd] followed by a phrase boundary.

[3] I abstract away from some aspects of the data by lumping all consonants together. Labov (1989) and Guy (1991a, 1994), amongst others, show that some consonants are more likely than others to induce deletion. Specifically, less sonorous consonants typically are more likely to result in deletion (i.e. more deletion in *best book* than in *best week*). Syllable structure constraints may also play a role (though cf. Labov 1997). For instance, Guy (1991a, 1997) points out that more deletion is observed before [l] than before [ɹ] (e.g. more deletion in *best luck* than *best rock*), which may be due to the fact that [tɹ-] is a possible onset cluster but [tl-] is not.

2.2 The analysis

The grammars for Chicano and Tejano type dialects are given in (6). The only difference between these two grammars is in the relative ranking of *Ct#V and *Ct##. In Chicano, *Ct#V ranks higher, corresponding to the fact that pre-vocalic position is associated with higher deletion rates in this dialect. The opposite holds for Tejano.

(6) Chicano: $\text{MAX}_{L1} \gg$ *Ct#C $\gg \text{MAX}_{L2} \gg$ *Ct#V $\gg \text{MAX}_{L3} \gg$ *Ct## $\gg \text{MAX}_{L4}$
Tejano: $\text{MAX}_{L1} \gg$ *Ct#C $\gg \text{MAX}_{L2} \gg$ *Ct## $\gg \text{MAX}_{L3} \gg$ *Ct#V $\gg \text{MAX}_{L4}$

Interspersed between the markedness constraints are lexically indexed versions of MAX. I follow Pater (2000; also Itô & Mester 1999) in assuming that a lexically indexed constraint only evaluates lexical items that are co-indexed with the constraint. I diverge from Pater, however, by proposing that lexical items do not have to be associated with a specific lexical class, but that they can vary in their class affiliation. Every time that a lexical item is submitted to the grammar for evaluation, it is assigned to one specific lexical class, determining which indexed constraint will evaluate it. Since a single lexical item can be assigned to different lexical classes on different evaluation occasions, it can be evaluated by different constraints, and this can result in variable output selection. The tableaux in (7) and (8) illustrate this by showing how an input like *best offer* will be evaluated under different indexations in Chicano and Tejano English.

(7) **Chicano**

/best$_{L2}$ offer/	MAX$_{L1}$	*Ct#C	MAX$_{L2}$	*Ct#V	MAX$_{L3}$	*Ct##	MAX$_{L4}$
☞best$_{L2}$ offer				*			
bes$_{L2}$ offer			*!				

/best$_{L3}$ offer/	MAX$_{L1}$	*Ct#C	MAX$_{L2}$	*Ct#V	MAX$_{L3}$	*Ct##	MAX$_{L4}$
best$_{L3}$ offer				*!			
☞ bes$_{L3}$ offer					*		

(8) **Tejano**

/best$_{L2}$ offer/	MAX$_{L1}$	*Ct#C	MAX$_{L2}$	*Ct##	MAX$_{L3}$	*Ct#V	MAX$_{L4}$
☞best$_{L2}$ offer						*	
bes$_{L2}$ offer			*!				

/best$_{L3}$ offer/	MAX$_{L1}$	*Ct#C	MAX$_{L2}$	*Ct##	MAX$_{L3}$	*Ct#V	MAX$_{L4}$
☞best$_{L3}$ offer						*	
bes$_{L3}$ offer					*!		

In Chicano, $\text{M}\small{\text{AX}}_{L1}$ and $\text{M}\small{\text{AX}}_{L2}$ outrank *Ct#V. If *best* is assigned to lexical classes L1 or L2 it will therefore be protected by a faithfulness constraint that ranks higher than the markedness constraint, and the non-deletion candidate will be selected as optimal. This is shown in the first tableau in (7). Since $\text{M}\small{\text{AX}}_{L3}$ and $\text{M}\small{\text{AX}}_{L4}$ rank below *Ct#V, assignment of *best* to L3 or L4 will result in deletion, as shown in the second tableau in (7). In Chicano, two out of four possible indexations result in deletion. In Tejano, however, *Ct#V is dominated by $\text{M}\small{\text{AX}}_{L1}$, $\text{M}\small{\text{AX}}_{L2}$ and $\text{M}\small{\text{AX}}_{L3}$. Three of the possible indexations therefore result in non-deletion, corresponding to the lower relative deletion rate in pre-vocalic position in Tejano than in Chicano. In a similar manner, we can determine the percentage of possible indexations that will result in deletion in each of the three contexts for these two dialects. The results of this calculation are given in (9). In both dialects, pre-consonantal position is associated with the highest likelihood of deletion. However, because of the relative ranking of *Ct#V and *Ct##, these dialects diverge with regard to the likelihood of deletion in pre-vocalic and phrase-final position. The relative contribution of the grammar (the ranking between the markedness constraints) to the likelihood of *t/d*-deletion is captured in this account.

(9) **Effects of different indexations in Chicano and Tejano**

Context	Indexations resulting in deletion		% deletion indexations	
	Chicano	Tejano	Chicano	Tejano
Pre-C	L2, L3, L4	L2, L3, L4	75	75
Pre-V	L3, L4	L4	50	25
Pre-##	L4	L3, L4	25	50

3. Usage frequency and *t/d*-deletion

As shown in §1, there is evidence that usage frequency also influences the likelihood of *t/d*-deletion, with deletion being more likely to apply to words that are used more frequently. Bybee's (2000) reanalysis of Santa Ana's (1991) Chicano corpus is the clearest illustration of this. However, Bybee's reanalysis of these data does not show whether the contribution of frequency and grammar are independent from each other, or whether there is a more complex interaction between these two factors. Concretely: does frequency result in higher deletion rates across all three of the grammatically defined contexts discussed in §2?

3.1 The independence of grammar and usage frequency

In order to determine whether the influence of usage frequency and the grammar is independent from each other, I conducted a small scale experiment. I selected 15 monomorphemic English words that end on [-st], and divided this list into a high frequency (*host*, *dust*, *fast*, *test*, *list*, *rest*, *best*, *last*, *most*) and low frequency (*crust*, *feast*, *yeast*, *mast*, *moist*, *nest*) group. Words in the high group have frequencies higher than the average frequency in Kučera and Francis (1967), and the low group has frequencies lower than this average. Each of these words were embedded into a sentence where it occurred as the last word in the sentence, followed by a consonant initial word, or followed by a vowel initial word. Six native English speakers (undergraduate students at the University of Michigan, all of whom grew up in southeastern Michigan) were recruited. After illustrating to the participants that word-final /t/ can sometimes be dropped in pronunciation, they were presented with the 45 sentences in randomized order in written form. They were asked to rate for each of the token words how likely they are to delete the /t/ in a casual speech situation. Rating was done on a 10 point scale where [1] meant that the /t/ is nearly always pronounced, and [10] that it is nearly never pronounced. The task that the participants performed is similar to a well-formedness rating: they rated each token for how well-formed a pronunciation without a [t] would sound. The scores for each participant were normalized so that a positive score on an item means that a participant is more likely than average to delete [t] on the specific item, and a negative score that he/she is less likely than average to delete [t]. The results are represented graphically in Figure 1. The results were averaged by item and submitted to an ANOVA with frequency (High, Low) and context (Pre-##, Pre-V, Pre-C) as factors. This analysis returned a significant result for frequency ($F(1,39) = 5.12$, $p < .03$) and context ($F(2,39) = 43.48$, $p < .001$), but not for the interaction between these two factors ($F(2,39) = 0.15$, $p = .86$). These results show that both grammar and frequency influence the likelihood of *t/d*-deletion. But importantly, the lack of an interaction between these two factors indicates their influence is independent from each other. *t/d* is more likely to delete from a high frequency word in each of the three conditions. In the rest of this section, I will show how the frequency of a lexical item can influence the likelihood of it being assigned to a specific lexical class, and hence the likelihood that *t/d*-deletion will apply to it.

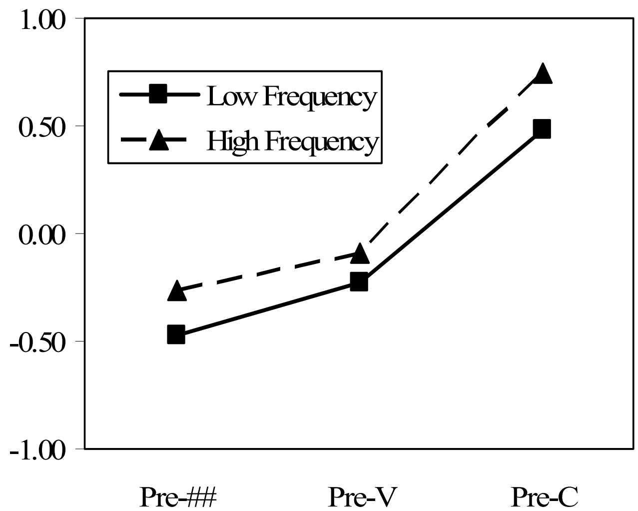

Figure 1: Average normalized ratings

3.2 Accounting for the influence of usage frequency on variation

As shown in §2, whether or not deletion applies to some form depends on the lexical class to which it is assigned. If the usage frequency of a lexical item can influence the likelihood of the lexical item being assigned to a specific lexical class, then usage frequency can therefore influence the likelihood of deletion applying to that item. I propose that each lexical item has a probability distribution function associated with it. The shape of these distribution functions is determined by the frequency of the lexical item. Lexical items with a frequency equal to the average lexical usage frequency have symmetric distributions. However, the higher the frequency of an item, the more its distribution function will be left skewed, and the less frequent an item the more its function will be right skewed. These distribution functions range over an area that is divided into equally sized regions corresponding to each of the lexical classes. Regions corresponding to lexical classes that are associated with higher ranking constraints appear to the left, and regions associated with lower ranked constraints appear to the right. When a lexical item is submitted to the grammar, a value is chosen randomly from the distribution defined by its distribution function, which determines the lexical class that the lexical item will be assigned to on this specific occasion.

This scenario is represented visually in Figure 2. In this figure, the distribution functions of a word with higher frequency (*absent*), average frequency (*suspect*) and lower frequency (*lift*) are represented. The distribution mass of the lower frequency *lift* is concentrated towards the left end of the range. Consequently, when a value is randomly selected from the area defined by *lift*'s distribution function, the likelihood is greater that a value will be selected that corresponds to lexical class L1, than a value that corresponds to L4. For more frequent word *absent*, the opposite is true. It is more likely that a value will be selected that corresponds to L4 than L1. The consequence of this is that a less frequent word like *lift* is more likely to be evaluated by a high ranking faithfulness constraint indexed to L1, while a more frequent word like *absent* is more likely to be evaluated by a low ranking faithfulness constraint indexed to L4. Since an infrequent word is more likely to be protected by a high ranking faithfulness constraint and a frequent word by a low ranking faithfulness constraint, it follows naturally that a frequent word is more likely to undergo deletion than an infrequent word.

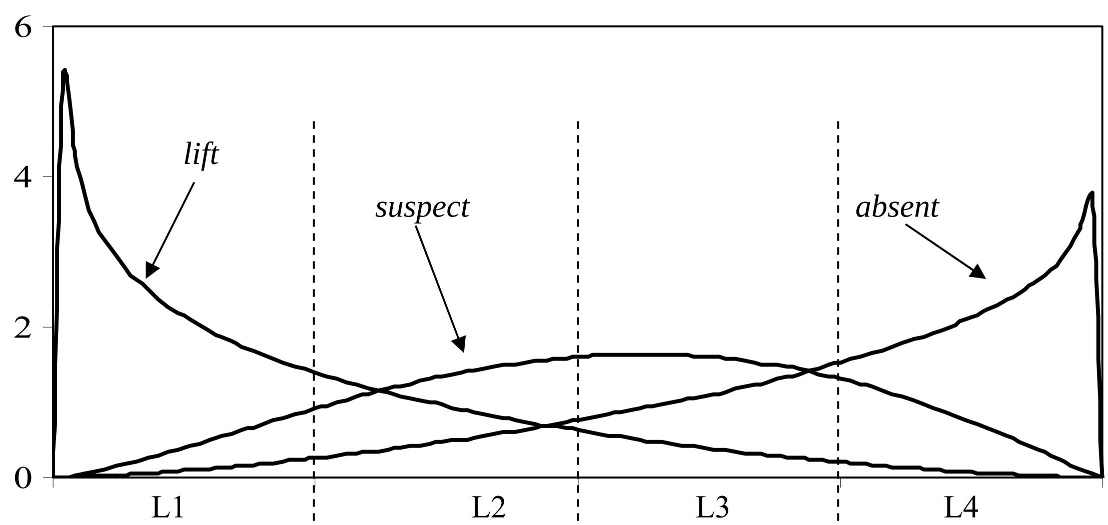

Figure 2: Lexical distribution functions

3.3 On the interaction between grammar and usage frequency

From this analysis, we can make two predictions about how grammar (the constraint ranking) and usage frequency (the lexical distribution functions) will interact. First, for any given individual word, irrespective of its usage frequency, most deletion will be observed in pre-consonantal position, and least in phrase-final position. This follows directly from the ranking between the three markedness constraints, and is hence the result of grammar. Secondly, given two lexical items lex_1 and lex_2, with lex_1 more frequent than lex_2, lex_1 will show higher deletion rates than lex_2 across all contexts. This follows from the fact that the more frequent lex_1 will be more likely to be evaluated by low ranking faithfulness constraints because of the shape of its distribution function. We do not currently have detailed enough information about *t/d*-deletion rate for any English dialect to test the validity of these two predictions. However, these are predictions that follow from the basic architecture of the grammatical model proposed here, and that hence provide an ideal opportunity for testing this model.

These predictions also imply that it is not very informative to compare the deletion rates of two lexical items. A highly frequent lexical item like *just* may have a higher deletion rate in the most conservative context (phrase-finally) than an infrequent lexical item like *bust* in the most liberal context (pre-consonantally). Interpreted out of context, this could lead to the impression that grammar (the phonologically defined contexts) do not really contribute to the likelihood of deletion. It is consequently very important to always look at the way in which a variable process influences a specific lexical item as a part of a larger system, and not just as an individual data point.

A final point that deserves mention is that grammar is still the primary decider of whether or not variation will be observed. Within the model developed here, the necessary conditions for variation to be observed are: (i) There must be some markedness constraint *M* whose violation can be avoided by violation of faithfulness constraint *F*. (ii) There must be at least two lexically indexed versions of *F*, F_{L1} and F_{L2}. (iii) F_{L1} and F_{L2} must rank on different sides of *M*. If these conditions are not met, then no variation will be observed at all, and usage frequency will have no influence on the output of the grammar. If these conditions are met, variation will be observed and only then will usage frequency be able to influence the way in which this variable process influences specific lexical items. To make this more concrete, I discuss two examples of non-variation in English.

First, an example of categorical non-application: Unlike German and Dutch, English tolerates voiced obstruents in coda position. If we assume that coda obstruent devoicing is motivated by a constraint *VOICEDOBS]$_\sigma$ (Itô & Mester 2003), then it follows that all lexically indexed versions of the faithfulness constraint IDENT[voice] must rank higher than *VOICEDOBS]$_\sigma$ in English. Consequently, it would not matter which version of IDENT[voice] evaluates a specific lexical item, any candidate that violates IDENT[voice] will always be worse that the faithful candidate. Not even a very frequent word that is very likely to be evaluated by low ranking IDENT[voice]$_{L4}$ will ever undergo devoicing. Usage frequency alone cannot force devoicing. This is illustrated in the tableau in (10) for the frequent word *and*. Since this word is high in frequency, it is more likely to be assigned to lexical class L4. The tableau here therefore represents the most likely scenario for the word *and*. Not even when it is assigned to L4, does the devoicing candidate beat the faithful candidate.

(10) **No variable obstruent devoicing**

/and$_{L4}$/	ID[voi]$_{L1}$	ID[voi]$_{L2}$	ID[voi]$_{L3}$	ID[voi]$_{L4}$	*VOICEDOBS]$_{\sigma}$
☞and$_{L4}$					*
ant$_{L4}$				*!	

Compulsory application of a process is handled similarly. Intra-morphemically English does not tolerate heterorganic [nasal + consonant] clusters. Every input that contains such a sequence must therefore be mapped unfaithfully onto an output where the nasal has assimilated in place to a following consonant. All lexically indexed versions of the faithfulness constraint IDENT[place] therefore have to rank lower than the markedness constraint AGREEPLACE$_{Morph}$ that dictates place agreement between a nasal and a following consonant within a morpheme. Consequently, it does not matter to which lexical class a lexical item is assigned, it will always undergo place assimilation. Not even a very infrequent lexical item could resist application of this process. This is shown in the tableau in (11), where the lexical item *ember* is evaluated. Under "richness of the base" any input must be mapped grammatically onto some output, and this tableau therefore shows what would happen if a child had (incorrectly) learned the underlying representation of *ember* as /ɛnbɚ/.

(11) **Compulsory place assimilation intra-morphemically**

/ɛnbɚ/	AGRPLACE$_{Morph}$	ID[place]$_{L1}$	ID[place]$_{L2}$	ID[place]$_{L3}$	ID[place]$_{L4}$
ɛnbɚ	*!				
☞ ɛmbɚ		*			

The model developed here is hence still a grammatical model of variation. Grammar determines whether or not variation will be observed. Grammar also influences the likelihood that a variable process will apply in different contexts. Usage frequency only becomes relevant once variation has been made possible by grammar.

4. Conclusion

In the literature on phonological variation, grammatical accounts and usage based accounts are often pitted against each other as opposing and incompatible. In the classical generative tradition, usage frequency and the ways in which it impacts linguistic behavior would have been considered as just this: part of performance/behavior and hence not in the domain of things for which phonology should account. In the usage-based literature, on the other hand, evidence that usage frequency influences variation is sometimes interpreted as showing that the complete process can be reduced to just frequency and that formal phonological grammar is not required. After showing how usage frequency influences *t/d*-deletion rate in Santa Ana's (1991) Chicano corpus Bybee (2000:73), for instance, concludes this shows that "there is no variable rule of *t/d*-deletion".

In this paper, I have shown that these two accounts of variation are not in principle incompatible with each other. It is possible to design grammatical models that simultaneously account for both the influences from formal grammar and the influences from usage frequency. This is not only possible, but ultimately necessary. The evidence showing that both of these factors influence variation is mounting, and shows that the language user has access to both kinds of information. His/her linguistic competence encompasses both of these factors.

Appendix A: Lexical distribution functions as instantiations of the *beta* distribution

In the model developed above, the lexical distribution functions have the following properties: (i) They have a finite range – this follows from the fact that the range of the functions must be exhaustively divided into regions corresponding to the different lexical classes. (ii) They can change shape. Specifically, they can change the area along their range where the mass of the distribution is concentrated (they can change skewness). In order to model the influence of the lexical distribution functions on variation, we therefore need a distribution function that has these properties. The *beta* distribuition (Evans *et al.* 2000; Gupta & Nadarajah 2004) meets this requirement. The probability density function of the *beta* distribution is given in (12). The range of this function is (0, 1), and α and β are shape parameters. The values of α and β determine the skewness of the distribution function as shown in the table in (13).

(12) **Probability density finction of the *beta* distribution**

$$f(x,\alpha,\beta)=\frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du}$$

(13) **The skewness of the *beta* distribution**

Shape parameters	*Shape of distribution*
$\alpha = \beta$	Symmetric
$\alpha < \beta$	Right skewed
$\alpha > \beta$	Left skewed

In the model developed above, I argued that usage frequency influences the shape of the lexical distribution functions. If the lexical distribution functions are interpreted as instantiations of the *beta* distribution, the usage frequency of a lexical item can be used to set the values of α and β for the distribution function of the specific lexical item, which will allow for frequency to determine the shape of the distribution function. In the rest of this appendix, I discuss a specific implementation of this proposal, in which I use the *Irvine Phontactic Online Dictionary* (IPhOD, www.iphod.com) as a corpus to model.

IPhOD is a freely available corpus consisting of 33,432 English words in phonetic transcription (using the CMU transcriptions, Weide 1994), each annotated with its corresponding frequency from Kučera and Francis (1967). I calculated the mean frequency of all the words in IPhOD (29.8), and then also selected all words that end on [-Ct] or [-Cd] (excluding past tense verbs) and determined their frequency in IPhOD. These frequency counts were multiplied by 10 before they were log transformed.[4] The log transformed frequency values were then used to set the values of α and β. The manner in which this was done is shown in the table in (14).[5]

(14) **Setting shape parameters using frequency information**

Frequency of item	*Skewness*	*Value of shape parameters* α	β
Frequent ($f > \mu$)	Left ($\alpha > \beta$)	$\log(\mu)$	$\log(\mu) - \log\lvert\mu - f\rvert$
Infrequent ($f < \mu$)	Right ($\alpha < \beta$)	$\log(\mu) - \log\lvert\mu - f\rvert$	$\log(\mu)$
Average ($f = \mu$)	Symmetric ($\alpha = \beta$)	$\log(\mu)$	$\log(\mu)$

(μ = mean frequency in IPhOD, and f = frequency of specific lexical item.
$\alpha \leq 0$ and $\beta \leq 0$ are not defined. Any such values are replaced with 0.01.)

The table in (15) gives the frequency values (multiplied by 10) for a representative sample of words from IPhOD, as well as the values of α and β for the

[4] Log transformation of frequency information is standard in psycholinguistic work. The multiplication by 10 is to avoid the problems caused by having to work with $\log_{10}(1) = 0$.

[5] The assumption made here in the setting of the shape parameters is that the behavior of lexical items is determined in relation to that of a lexical item with an average frequency. Whichever of α or β has to be larger, is set as the log of the mean frequency of lexical items in the corpus. The other parameter is set as follows: The absolute size of the difference of the mean frequency and the the frequency of the specific lexical item is determined, and the log of this absolute difference is subtracted from the log of the mean. The value of the parameter that has to be highest is hence the log of the mean. The value of the remaining parameter is smaller than the log of the mean, and the amount that it is smaller depends on how much the frequency of the specific lexical item differs from the mean frequency.

distribution function of each of these words. Figure 3 represents the distribution functions of these words, imposed over the grammar developed above.

(15) **Frequencies and shape parameters for selected words**

μ = 297.9	*assist*	*fist*	*belt*
Frequency ($\times$ *10*)	260	280	300
α	log(μ) - log|μ - f| 0.90	log(μ) - log|μ - f| 1.22	log(μ) 2.48
β	log(μ) 2.48	log(μ) 2.48	log(μ) - log|μ - f| 2.15

μ = 297.9	*lieutenant*	*absent*
Frequency ($\times$ *10*)	320	340
α	log(μ) 2.48	log(μ) 2.48
β	log(μ) - log|μ - f| 1.13	log(μ) - log|μ - f| 0.85

From Figure 3, it is clear that deletion is more likely to apply to words that appear more frequently – since the mass of their distributions is centered to the right end of the range, and therefore over the parts of the range that correspond to lower ranked faithfulness constraints. However, now that the distribution functions have been implemented as instantiations of the *beta* distribution, it is possible to go beyond mere informal observations like this. If the values of α and β are known, then the probability density function of the *beta* distribution, given in (12) above, can be used to calculate the exact portion of the distribution mass of the function that covers a specific part of the range. Concretely, since there are four lexical classes in the grammar under discussion here, the (0, 1) range of the lexical distribution functions are divided into four equally sized ranges, (0, .25], (.25, .50], (.50, .75], (.75, 1). With α and β known, we can calculate the proportion of the distribution mass that covers each of these four ranges, and hence for every lexical item the likelihood of it being assigned to each of the four lexical classes, and hence the likelihood of it being subject to deletion in each of the three contexts.

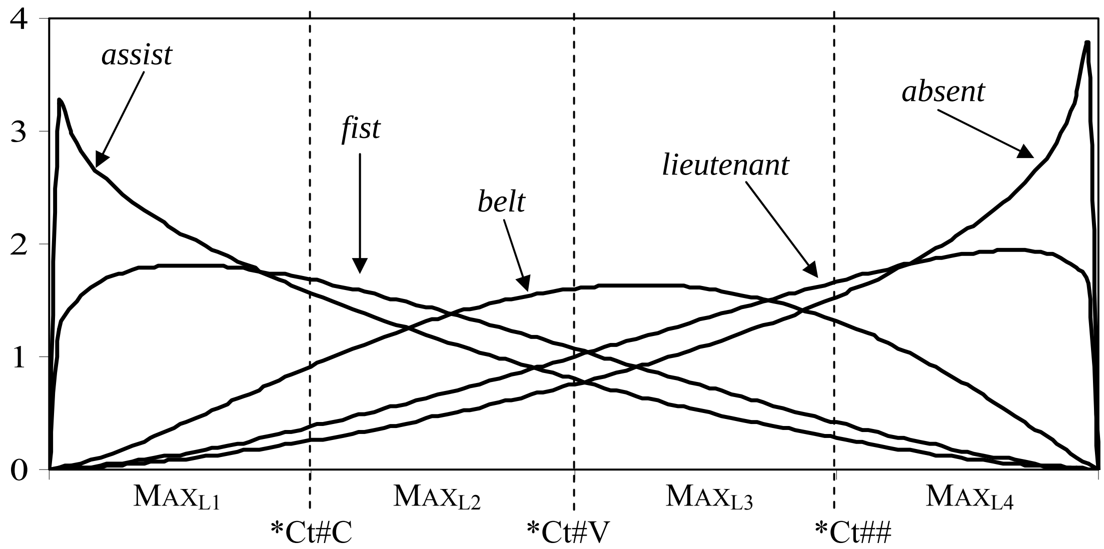

Figure 3: Lexical distribution functions for some words from IPhOD

Using this interpretation of the lexical distribution functions, I determined the value of α and β for each of the (non past-tense) lexical items in IPhOD that end on [-Ct] or [-Cd], as explained above. Using the probability density function of the *beta* distribution, I calculated the proportion of the distribution mass of the lexical distribution functions that covers each of the four regions on the range of the functions that correspond to lexical classes L1, L2, L3, and L4. Once these proportions are known, the predicted likelihood of deletion for any lexical item in each of the three contexts (pre-consonantal, pre-vocalic and phrase-final) can be calculated. The table in (17) shows the results of these calculations for the lexical items whose distribution functions are represented in Figure 3. In this table, I make the assumption that each lexical item appears in each of the three contexts at its frequency listed in IPhOD (× 10, as explained above). The last line in this table sums the results for all the (non past-tense) lexical items that end on [-Ct] or [-Cd].

Inspection of the table in (17) will confirm that this implementation of the lexical distribution functions do generate expected deletion rates that are in agreement with the predictions outlined in §3.3 above. For any given word, deletion rates are expected to be the highest in pre-consonantal position, intermediate in pre-vocalic position, and lowest in phrase-final position. Also, given two words that differ in frequency, the expected deletion rate of the more frequent word is higher than that of the less frequent word in each of the three contexts. In general, interpreting lexical distribution functions as instantiations of the *beta* distribution therefore seems to work. However, if we compare the expected deletion rates in the three contexts for the corpus as a whole with the actually observed deletion rates in Santa Ana's (1991) Chicano corpus, it is clear that the fit between expected and observed frequency is not perfect. This comparison is shown in (16).[6]

(16) **Expected and observed deletion rates in Chicano English**

	Pre-Consonantal	*Pre-Vocalic*	*Phrase-Final*
Observed	62%	45%	37%
Expected	83%	80%	77%

The expected relative deletion rates agree with the observed rates in Chicano English, but the fit in terms of absolute frequency is less good. Additionally, in the current model, lexical usage frequency is the sole determiner of the shape of the lexical distribution functions. If we assume that all dialects of English have roughly the same lexicon with the same usage frequencies for different lexical items, it follows that the lexical distribution functions will be the same for all dialects of English. Any English dialect with the same grammar (constraint ranking) as Chicano English will therefore also have the same expected deletion rates. And we know that this is not the case. There are several dialects with the same relative deletion rates between the contexts, but with very different absolute deletion rates (see Coetzee 2004: Chapter 5, and Coetzee & Pater 2008 for examples).

It seems to be that the implementation of the lexical distribution functions as instantiations of the *beta* distribution is not very successful. However, there are reasons to be less pessimistic. First, the way in which the relation between frequency and the shape parameters, α and β, is implemented here is but one of many different ways in which this could be done. It is possible that a different relation between frequency, and α and β will result in a better fit between expected and observed deletion rates.

Another option to explore is to add extra parameters to the current model. For instance, a scaling factor could be added to the function that relates frequency to α and β. The values of α and β are currently determined as follows: Whichever of these two parameters needs to be largest is set equal to $\log(\mu)$, and the other is set equal to $\log(\mu) - \log|\mu - f|$. One possible implementation of a scaling factor would be to multiply $\log|\mu - f|$ by some constant value. If this scaling factor is set smaller than 1, the result will be that the values of α and β will differ less from each other. This, in turn, will imply that the lexical distribution functions will all on average be less skewed, so that frequent and infrequent words will differ less from each other. The model described above where $\log|\mu - f|$ was not scaled would then be the default system where the scaling factor is set to 1.

[6] The Excel file that was used to run this simulation, as well as the simulation reported in (18) below, is available from www.umich.edu/~coetzee/Frequency_and_Variation.

(17) **Predicted deletion rates for a selection of lexical items from IPhOD**

Word	Probability of being assigned to lexical class				Expected deletion rates								
					Pre-Consonantal			Pre-Vocalic			Phrase-Final		
	L1	L2	L3	L4	% deleted	# deleted	# retained	% deleted	# deleted	# retained	% deleted	# deleted	# retained
assist	.55	.29	.13	.03	45%	116	144	16%	41	219	3%	7	253
fist	.42	.35	.19	.04	58%	162	118	23%	64	216	4%	12	268
belt	.10	.33	.39	.18	90%	269	31	57%	170	130	18%	53	247
lieutenant	.04	.17	.33	.46	96%	308	12	79%	253	67	46%	146	174
absent	.03	.12	.28	.57	97%	331	9	85%	290	50	57%	195	145
Corpus total					83%	307,911	63,019	80%	296,198	74,732	77%	286,013	84,917

Comparing the observed and the expected deletion rates in (16) shows that the deletion rates are on average too high. This is most likely due to highly frequent words having distribution functions that are too severely left-skewed so that they are predicted to undergo deletion more often than actually observed. The influence of frequency could be ameliorated by manipulating the scaling factor. To test this, I repeated the simulation, this time using a scaling factor of 0.5. The value of the higher of α and β was still set to $\log(\mu)$, but the value of the lower of two was set to $\log(\mu) - 0.5(\log|\mu - f|)$. The results of this new simulation are given in the table in (18). As expected, applying this scaling factor does bring the expected and observed deletion rates closer together.

(18) **Expected deletion rates in Chicano English with a scaling factor of 0.5**

	Pre-Consonantal	*Pre-Vocalic*	*Phrase-Final*
Observed	62%	45%	37%
Expected	90%	73%	46%

Although this results in a better fit between the observed and predicted deletion rates, the fit is still not very good. However, this is but one of many different ways in which a scaling factor could be built into the model. Rather than proposing that this is the way to apply scaling, the purpose of this illustration is to show how the introduction of additional parameters such a scaling factor could influence the predictions of the model. Further research would be required to determine what the best way is in which to implement a scaling factor, or even whether this would be the best way in which to augment the model. However, this does seem to be a way in which the differences between dialects, and also the differences between speech styles could be accounted for. Different dialects could have different scaling functions. Similarly, there could be different scaling functions for different speech styles which could account for the fact that deletion is more likely to happen in faster and casual speech than in careful and slow speech (Browman & Goldstein 1990).

The important point to take away is that the interpretation of the lexical distribution functions as instantiations of the *beta* distribution meets the basic requirements for the variation model proposed in this paper. This model makes specific predictions about how the usage frequency of different words will impact their participation in variable processes (see §3.3 above for details). These basic predictions are borne out by the interpretation of the lexical distribution functions given here. In order for the predictions of the model to fit the data better, further refinement of the basic model is necessary, and as shown just above, also possible.

Appendix B: The acquisition path of lexical distribution functions

In an OT model without lexical distribution functions, the learner has to learn only the grammar, i.e. the constraint ranking. In the model proposed here, the task of the learner is more difficult. In addition to learning the grammar, he/she also has to keep track of lexical usage frequencies, and update the lexical distribution functions of lexical items in accordance with these frequencies. Pater (2005, see also Pater & Coetzee 2005, and Coetzee 2008) proposes an extension to the learning algorithm of Tesar and Smolensky (1998) that enables the algorithm to learn a grammar with lexically indexed faithfulness constraints. Since an algorithm exists that can learn the type of OT grammar that I assume in the model above, I do not consider the acquisition of the grammar further here. The rest of this appendix is dedicated to investigating the acquisition path of lexical distribution functions.

The actual acquisition of the lexical distribution functions is a trivial task. If we assume that the learner comes equipped with the general formula for the *beta* distribution and with knowledge of how to determine the values of the shape parameters α and β based on usage frequency, then no real acquisition is required. All that the learner has to do is to keep track of usage frequencies, and to adjust the values of α and β according to the changing values of the usage frequencies. In the rest of this appendix, I show how lexical distribution functions change over time as a function of increased exposure, and what influence this has on the variation patterns.

In (14) above, I proposed that the values of α and β are set in relation to the mean frequency of lexical items. A question that needs answering is what the total size of the corpus is over which the language user calculates usage frequencies. Since the language user is exposed to language use on an ongoing basis, it is in principle possible that the corpus size has no upper bound but that it keeps expanding throughout the lifespan of the language user. This would entail that the language user stores a record of every lexical item that he/she is exposed to throughout his/her life. Although we know that language users have an amazing ability to store a large number of detailed exemplars, there is also evidence that these exemplar memories fade over time (Goldinger 1996, 1997, 1998). It is therefore unlikely that the corpus used for frequency calculations keeps on expanding indefinitely. The corpus more likely has some upper size limit, and once this limit has been reached, older entries fade from the corpus as newer entries are added through exposure. It also seems reasonable to assume that the upper limit for the corpus size might be subject to a developmental timeline. During the earlier stages of language development, the upper limit might be lower so that entries start fading from the corpus sooner. As the child matures, the upper limit of the corpus size increases, so that older entries are retained longer in the corpus.

Based on the reasoning in the previous paragraph I make these two assumptions: (i) There is an upper limit to the size of the corpus over which lexical frequencies are calculated. (ii) The upper limit of the corpus size increases with mental development – i.e. the corpus over which frequencies are calculated is smaller for children still developing in their language abilities. The rest of this appendix is dedicated to exploring the implications of these two assumptions. If they are correct, what do we expect to observe in terms of the interaction between lexical frequency and variation during the period of linguistic development?

Per the proposal in the previous appendix, the values of the shape parameters α and β are equal to either log(μ), or log(μ) - log|μ - f|. How does an increase in corpus size affect these values? As corpus size increases the values of both f and μ also increase, and further more, they increase at an equal rate – if the total corpus size increases by a factor of *x*, then both μ and f also increase by a factor of *x*. From this follows the following: (i) The values of log(μ) and log|μ - f| also steadily increase with increasing corpus size. (ii) The value of log(μ) - log|μ - f| remains constant. (iii) The value of one of the shape parameters remains constant, while the value of the other parameter steadily increases. (iv) The difference between the shape parameters steadily increases. This is illustrated in the tables in (19) and (20) for a mini lexicon with four words, where $Word_4$ appears four times as often as $Word_1$, $Word_3$ three times as often, and $Word_2$ twice as often. Figure 4 plots the difference between α and β for each of the four words in this mini lexicon as a function of the increase in corpus size. A negative difference implies that α is smaller than β, and hence that the distribution function will be right skewed (infrequent word), and the opposite for a positive difference. This figure shows that the difference between α and β steadily increases as the size of the corpus increases. This will have the effect that the skewness of the distribution functions will also increase with an increase over time. A prediction is therefore that the influence of frequency will become more extreme over time – frequent words will participate more and more in variable processes over time, and infrequent words will resist participation in the variable process more and more. Once the maximum size of the corpus has been reached, this increased influence of frequency will start to taper off.

(19) **Mini lexicon at 10 consecutive time points**

Time	*$Word_1$*	*$Word_2$*	*$Word_3$*	*$Word_4$*	*μ*	*Corpus Size*
1	10	20	30	40	25	100
2	20	40	60	80	50	200
3	40	80	120	160	100	400
4	80	160	240	320	200	800
5	160	320	480	640	400	1600
6	320	640	960	1280	800	3200
7	640	1280	1920	2560	1600	6400
8	1280	2560	3840	5120	3200	12800
9	2560	5120	7680	10240	6400	25600
10	5120	10240	15360	20480	12800	51200

(20) **Parameter settings for words in mini lexicon**

Times	$Word_1$			$Word_2$		
	α	β	α - β	α	β	α - β
1	0.22	1.40	-1.18	0.70	1.40	-0.70
2	0.22	1.70	-1.48	0.70	1.70	-1.00
3	0.22	2.00	-1.78	0.70	2.00	-1.30
4	0.22	2.30	-2.08	0.70	2.30	-1.60
5	0.22	2.60	-2.38	0.70	2.60	-1.90
6	0.22	2.90	-2.68	0.70	2.90	-2.20
7	0.22	3.20	-2.98	0.70	3.20	-2.51
8	0.22	3.51	-3.28	0.70	3.51	-2.81
9	0.22	3.81	-3.58	0.70	3.81	-3.11
10	0.22	4.11	-3.89	0.70	4.11	-3.41

Times	$Word_3$			$Word_4$		
	α	β	α - β	α	β	α - β
1	1.40	0.70	0.70	1.40	0.22	1.18
2	1.70	0.70	1.00	1.70	0.22	1.48
3	2.00	0.70	1.30	2.00	0.22	1.78
4	2.30	0.70	1.60	2.30	0.22	2.08
5	2.60	0.70	1.90	2.60	0.22	2.38
6	2.90	0.70	2.20	2.90	0.22	2.68
7	3.20	0.70	2.51	3.20	0.22	2.98
8	3.51	0.70	2.81	3.51	0.22	3.28
9	3.81	0.70	3.11	3.81	0.22	3.58
10	4.11	0.70	3.41	4.11	0.22	3.89

In order to get a more realistic picture of how an increase in corpus size influences the shape of lexical distribution functions and the application of a variable process, I used the IPhOD corpus to simulate three different developmental stages. IPhOD gives the frequency per million (Kučera & Francis 1967) for each word contained in the corpus. The expected frequency of each word in the corpus per *x* million exposures can therefore be calculated by multiplying the frequency listed in IPhOD by *x*. I created three versions of the corpus that correspond to the developmental stage when the corpus over which frequencies are calculated has a size of one million (untransformed IPhOD corpus), when it has a size of two million (multiplying IPhOD frequency of each word by two), and when it has a size of ten million (multiplying IPhOD frequency of each word by ten). I then calculated the values of the shape parameters α and β at each of these developmental stages for four of the words in the corpus that end on [-Ct], using the method explained in the previous appendix. The words were selected so that two appeared more frequently than average (*lieutenant*, *absent*) and two less frequently than average (*assist*, *fist*). The table in (21) lists the expected frequencies and the values of the shape parameters for each of these words at the three developmental stages. The table in (22) lists the predicted deletion rates for these four words in each of the three phonologically defined contexts discussed in this paper, at the three developmental stages. Finally, Figure 5 shows the lexical distribution functions for these four words at the three developmental stages modeled here.[7]

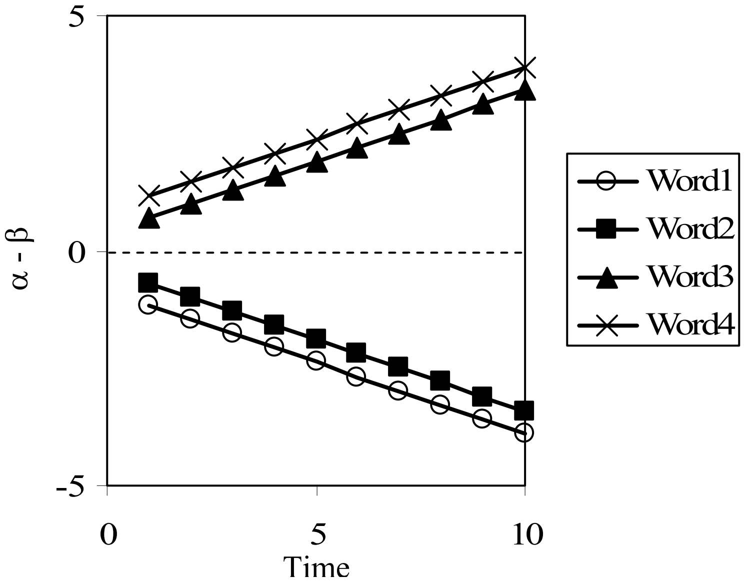

Figure 4: α- β for the words in the mini-lexicon

(21) **Frequencies and shape parameters at three different corpus sizes**

	1 million				*2 million*				*10 million*			
	f	*μ*	*α*	*β*	*f*	*μ*	*α*	*β*	*f*	*μ*	*α*	*β*
assist	26	29.8	0.90	1.47	52	59.6	0.90	1.78	260	297.9	0.90	2.47
fist	28	29.8	1.22	1.47	56	59.6	1.22	1.78	280	297.9	1.22	2.47
lieutenant	32	29.8	1.47	1.13	64	59.6	1.78	1.13	320	297.9	2.47	1.13
absent	34	29.8	1.47	0.85	68	59.6	1.78	0.85	340	297.9	2.47	0.85

In the table in (22), it is once again clear that this model does indeed predict most deletion for all words in pre-consonantal position and least in phrase-final position. This is true for each word in all three of the different developmental stages modeled here. This table also shows again how frequency influences deletion – in each of the three developmental stages, the expected deletion rate for more frequent words are higher than that for less frequent words. However, what is most remarkable in this table is how the expected deletion rates of a single word changes depending on the size of the corpus. For the two words that appear less often than average (*assist*, *fist*), deletion rates fall steadily

[7] The Excel file used in this simulation is available at www.umich.edu/~coetzee/Frequency_and_Variation.

across all three contexts as the size of the corpus increases. The infrequent words therefore resist deletion more and more, and slowly approach the point of categorical non-deletion. For the two words that have a higher than average frequency (*lieutenant*, *absent*), the opposite is true. As corpus size increases deletion rates increase for these words across all three contexts. Investigation of the lexical distribution functions of these words in Figure 5 makes the reason for these trends clear. The skewness of the distribution functions increases as the size of the corpus increases. The distribution mass of the lexical distribution functions of infrequent lexical items becomes more concentrated towards the left end of the range in later developmental stages. As a consequence, these words become more likely to be assigned to lexical classes indexed to high ranking faithfulness constraints. The opposite holds for high frequency lexical items.

(22) **Predicted deletion rates at three different corpus sizes**

	1 million			*2 million*			*10 million*		
	Pre-C	*Pre-V*	*Pre-##*	*Pre-C*	*Pre-V*	*Pre-##*	*Pre-C*	*Pre-V*	*Pre-##*
assist	.61	.33	.11	.55	.26	.07	.45	.16	.03
fist	.73	.43	.16	.68	.36	.11	.58	.23	.04
lieutenant	.85	.60	.30	.90	.67	.35	.96	.79	.46
absent	.89	.69	.41	.93	.75	.47	.97	.85	.57

The model of the interaction between lexical frequency and variation proposed in this paper therefore makes a prediction about how variation will affect lexical items at different developmental stages. Earlier in the developmental progression the corpus over which frequencies are calculated is smaller. As a consequence, the lexical distribution functions are less skewed at these earlier stages, so that the rate of application of a variable process will be closer to chance – i.e. an individual lexical item will vary more between alternative pronunciations. As the size of the corpus over which frequencies are calculated increases, the skewness of the lexical distribution functions increases. This results in lexical items patterning closer to the categorical endpoints of application or non-application – for a single lexical item, the preference for one specific variant over the alternative becomes stronger. In other words, individual lexical items are predicted to be pronounced with more variation at earlier developmental stages than at later developmental stages. Although we do not have longitudinal data about the application of specific variable processes during different developmental stages, it is true that the utterances of children who are still acquiring the grammar and lexicon of their language are generally more variable than that of adults. This prediction therefore seems to be correct.

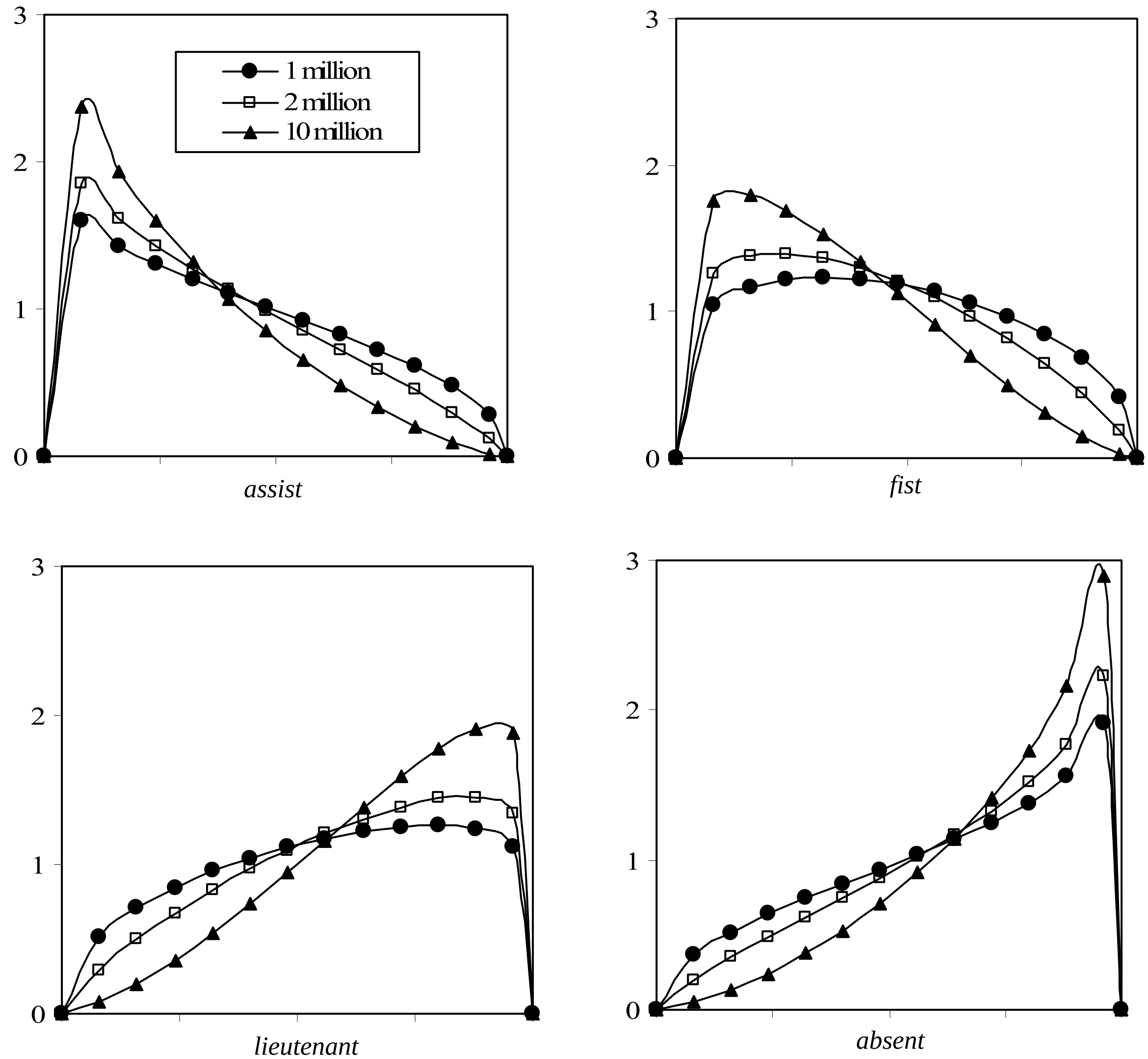

Figure 5: Lexical distribution functions at three different corpus sizes

These simulations also show the need for placing an upper limit on the size of the corpus over which frequencies are calculated. If this corpus just kept on expanding throughout the lifespan of the language user, then variation will keep on decreasing and eventually reach the point where no variation is observed – when the lexical distribution functions are so skewed that frequent words in effect are always assigned to the lexical class indexed to the lowest ranked faithfulness constraint, and infrequent words to the lexical class indexed to the highest ranked faithfulness constraint.

A last remark that needs to be made to about this model, is that it could be extended to account for some examples of language change. If some lexical item is extremely frequent, its lexical distribution function could be so heavily left skewed that it is nearly always assigned to the lexical class indexed to the lowest ranked faithfulness constraint. It will hence nearly always undergo the variable process. Continuing with the *t/d*-deletion example, such a lexical item will nearly always undergo deletion. A child acquiring the language might therefore never, or only very rarely, hear the word with its final *t/d* intact. It is therefore possible that the child might, due to lack of evidence for the presence of an underlying *t/d*, learn the lexical entry of the specific lexical item without the underlying *t/d* that is still present in the underlying form of his/her parents.

References

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

Anttila, Arto. 2002a. Morphologically Conditioned Phonological Alternations. *Natural Language and Linguistic Theory*, 20:1-42.

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

Bayley, Robert. 1997. Variation in Tejano English: evidence for variable Lexical Phonology. In C. Bernstein, T. Nunnaly, and R. Sabino, eds. *Language Variety in the South Revisited.* Tuscaloosa: University of Alabama Press.

Boersma, Paul. 1998. *Functional Phonology: Formalizing the Interaction Between Articulatory and Perceptual Drives*. The Hague: Holland Academic Graphics. [Doctoral dissertation, University of Amsterdam.]

Boersma, Paul and Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry*, 32:45-86.

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

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

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

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

Coetzee, Andries W. 2004. *What It Means to Be a Loser: Non-Optimal Candidates in Optimality Theory*. Ph.D. dissertation, University of Massachusetts Amherst.

Coetzee, Andries W. 2006. Variation as accessing "non-optimal" candidate. *Phonology*, 23:337-385.

Coetzee, Andries W. 2008. Grammaticality and ungrammaticality in phonology. *Language*, 84(2).

Coetzee, Andries W. and Joe Pater. 2008. *The Place of Variation in Phonological Theory*. Ms. University of Michigan and University of Massachusetts. [ROA #946.]

Donnegan, Patricia Jane and David Stampe. 1979. The study of Natural Phonology. In Daniel A. Dinnsen, ed. *Current Approaches to Phonological Theory.* Bloomington: Indiana University Press. p. 126-173.

Evans, Merran, Nicholas Hastings, and Brian Peacock. 2000. *Statistical Distributions. Third Edition*. New York: John Wiley & Sons.

Francis, W. Nelson and Henry Kučera. 1982. *Frequency Analysis of English Usage*. Boston: Houghton Mifflin.

Goldinger, Stephan D. 1996. Words and voices: episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22:1166-1183.

Goldinger, Stephan D. 1997. Words and voices: perception and production in an episodic lexicon. In Keith Johnson and John W. Mullenix, eds. *Talker Variability in Speech Processing*. London: Academic Press.

Goldinger, Stephen D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2):251-279.

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

Guy, Gregory R. 1991a. Contextual conditioning in variable lexical phonology. *Language Variation and Change*, 3:223-239.

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

Guy, Gregory R. 1994. The phonology of variation. In *CLS 30: Papers from the 30th Regional Meeting of the Chicago Linguistic Society. Volume 2: The Parasession on Variation in Linguistic Theory*, eds. Katharine Beals *et al.*, 133-149. Chicago: Chicago Linguistic Society.

Guy, Gregory R. 1997. Competence, performance, and the generative grammar of variation. In Frans Hinskens, Roeland van Hout and Leo Wetzels, eds. *Variation, Change and Phonological Theory*. Amsterdam: John Benjamins. p. 125-143.

Hayes, Bruce and May Abad. 1989. Reduplication and syllabification in Ilokano. *Lingua*, 77:331-374.

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

Itô, Junko and Armin Mester. 1999. The structure of the phonological lexicon. In Tsujimura Natsuko, ed. *The Handbook of Japanese Linguistics*. Malden: Blackwell. p. 62-100.

Itô, Junko and Armin Mester. 2003. On the sources of opacity in OT: coda processes in German. In Caroline Féry and Ruben van de Vijver, eds. *The Syllable in Optimality Theory*. Cambridge: Cambridge University Press. p. 271-303.

Kaisse, Ellen M., and Patricia A. Shaw. 1985. On the theory of Lexical phonology. *Phonology Yearbook*, 2:10-30.

Kiparsky, Paul. 1982. Lexical phonology and porphology. In In-Seok Yang, ed. *Linguistics in the Morning Calm*. Seoul: Hanshin. p. 1-91.

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

Kučera, Henry and W. Nelson Francis. 1967. *Computational Analysis of Present-Day American English*. Providence: Brown University Press.

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. 1997. Resyllabification. In Frans Hinskens, Roeland van Hout and Leo Wetzels, eds. *Variation, Change and Phonological Theory*. Amsterdam: John Benjamins. p. 145-179.

Labov, William. 2004. Quantitative analysis of linguistic variation. In Ulrich Ammon, Norbert Dittmer, Klaus J. Mattheier, and Peter Trudgill, eds. *Sociolinguistics: An International Handbook of the Science of Language and Society. 2nd edition*. Berlin: De Gruyter. p. 6-21.

Pater, Joe. 2000. Non-uniformity in English secondary stress: the role of ranked and lexically specific constraints. *Phonology*, 17:237-274.

Pater, Joe. 2005. Learning a stratified grammar. In Alejna Brugos, Manuella R. Clark-Cotton, and Seungwan Ha, eds. *Proceedings of the 29th Boston University Conference on Language Development*. Somerville: Cascadilla Press. p. 482-492.

Pater, Joe and Andries W. Coetzee. 2005. Lexically specific constraints: gradience, learnability, and perception. In *Proceedings of the Korea International Conference on Phonology*. Seoul: The Phonology-Morphology Circle of Korea. p. 85-119.

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.

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.

Santa Ana, Otto. 1992. Chicano English evidence for the exponential hypothesis: a variable rule pervades lexical phonology. *Language Change & Variation*, 4: 275-288.

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

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

Tesar, Bruce and Paul Smolensky. 1998. Learnability in Optimality Theory. *Linguistic Inquiry*, 29:229-268.

Weide, Robert L. 1994. *CMU Pronouncing Dictionary*. (www.speech.cs.cmu.edu/cgi-bin/cmudict.)

Department of Linguistics
University of Michigan
440 Lorch Hall
611 Tappan Street
Ann Arbor, MI 48109-1220

coetzee@umich.edu