This version will appear in the proceedings of NELS 38.
The "long version" is identical, except for the addition
of two appendices.

February 2008.

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, MAX_{L1} and MAX_{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 MAX_{L3} and MAX_{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 MAX_{L1}, MAX_{L2} and MAX_{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.

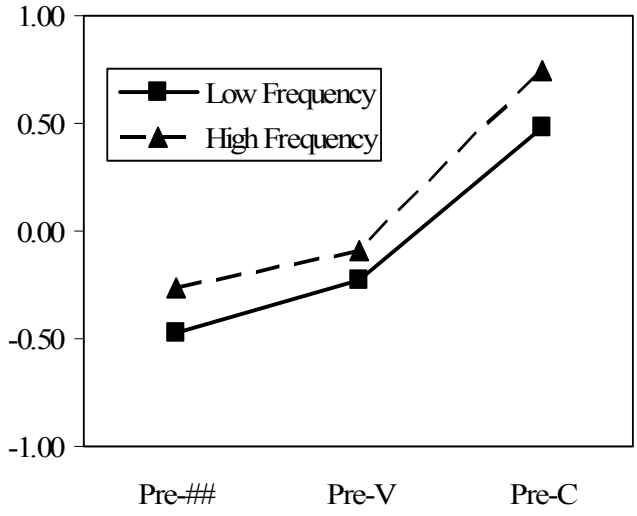

Figure 1: Average normalized ratings

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. [4]

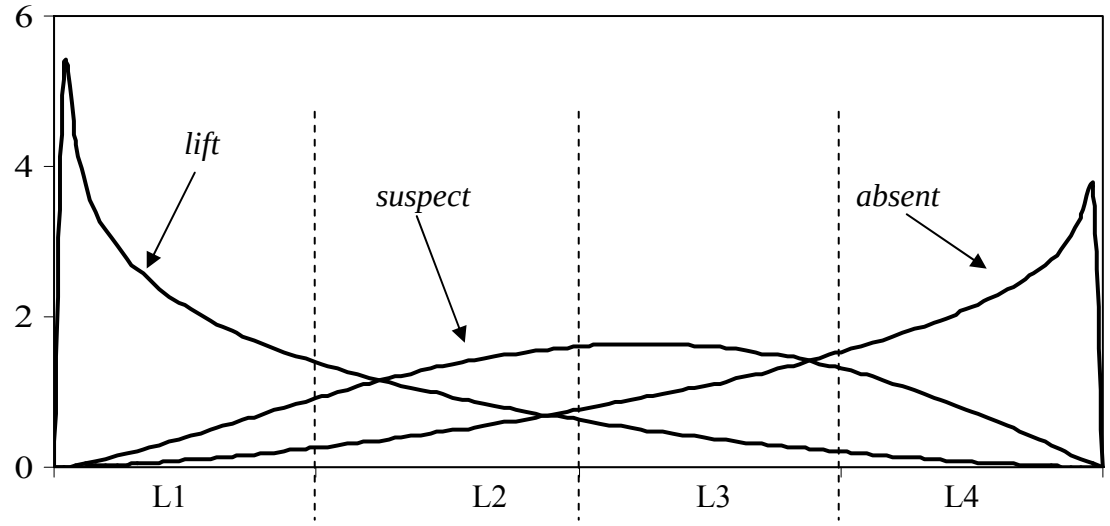

Figure 2: Lexical distribution functions

[4] A somewhat longer version of this paper is available on the Rutgers Optimality Archive. The longer version differs from this one only in the addition of two appendices. The first appendix interprets the lexical distribution functions as instantiations of the *beta* distribution (Evans *et al.* 2000; Gupta & Nadarajah 2004). The second appendix models the acquisition path of 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.

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.

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.

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.

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.

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