Please cite this paper as follows:

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

An integrated grammatical/non-grammatical model of phonological variation[1]

Andries W. Coetzee
University of Michigan
coetzee@umich.edu

Abstract

Phonological variation is conditioned by both grammatical and non-grammatical factors. Nearly all models of variation in phonological theory are exclusively grammatical, and do not account for the influence of non-grammatical factors at all. In this paper, I argue that an adequate theory of phonological variation should account for both grammatical and non-grammatical influences on variation, and I develop a noisy Harmonic Grammar model of variation that can do this. The constraints used in Harmonic Grammar are ordinary grammatical constraints (as in Optimality Theory), and the influence of grammar on variation follows from this. The influence of non-grammatical factors is accounted for by scaling the weight of faithfulness constraints up or down, resulting in less or more unfaithfulness.

1. Introduction

In early generative phonology, variation was regarded as not a core part of the phonological grammar: it was introduced late in phonological derivation (at the boundary between phonology and phonetic implementation), and was conditioned by non-grammatical factors. In Lexical Phonology (Kiparsky 1982), for instance, it was assumed that the core (lexical) phonological rules apply obligatorily, and that the only rules that can apply optionally are those mediating the interface between phonology and phonetic implementation (the post-lexical rules) (Kaisse & Shaw 1985:6; Kiparsky 1985:86).[2] Additionally, the lexical rules could only be conditioned by grammatical factors (morphology, phonology, etc.), while the post-lexical rules could be subject to external factors such as rate of speech (Kaisse & Shaw 1985:6). Although the influence of such external factors was acknowledged, it was considered to fall outside the domain of grammar, and the grammatical models developed in this tradition therefore had no formal mechanism for incorporating the influence of these

1 The research for this paper was partially supported by a research grant from the Horace Rackham Graduate School at the University of Michigan. This research has benefited greatly from discussion with especially Joe Pater, but also many other people (including John McCarthy, John Kingston, Pam Beddor, Kevin McGowan, Ania Musial, Anthony Brasher, and Jon Yip), and from audience feedback (MCWOP 11, NELS 38, NAPhC 5, Georgetown University, University of Michigan, University of Massachusetts).

2 The distinction made between obligatory rules and optional processes in Natural Phonology (Donnegan & Stampe 1979:145) is another example of how variation was pushed to the outer edges of phonological grammar.

factors. Additionally, because variation was considered at best a non-core part of phonological grammar, it received very little attention in this research tradition.

At the same time as these developments in theoretical phonology, variationist sociolinguistics in the Labovian tradition (Labov 1969) was also flourishing. In this tradition, variation was considered a core part of phonology. The architecture of the phonological component of grammar was therefore envisioned differently. Since all of phonological grammar can be variable, the distinction between *obligatory* lexical rules and *optional* post-lexical rules was considered to be a false distinction.[3] The distinction between grammatical and non-grammatical influences on the application of a variable process was also blurred. Non-grammatical factors could be incorporated into the structural description of a rule in exactly the same way as more traditional grammatical factors (Bayley 2002:126). Phonological grammar therefore included reference to non-grammatical factors, and, in fact, there was no formal difference between these two kinds of factors.

In spite of the fact that the variationist approach dominated sociolinguistic research during the 1970's and 1980's, theoretical phonology barely took note of the developments in this parallel field. It has only been since the early 1990's that variation began to occupy a more central place in the research focus of theoretical phonology, and that the theoretical phonology research tradition started to take note of the decades of research in variationist sociolinguistics – spearheaded by researchers such as Reynolds (1994) and Anttila (1997, 2002); see also Boersma (1997), Boersma and Hayes (2001), and Coetzee (2004, 2006, *to appear*). See Coetzee and Pater (*to appear*) for a review of the factors that lead to this change in how theoretical phonology views variation.

These new approaches to variation in theoretical phonology are very successful at accounting for variation. They not only account for where variation should be possible, but also for the frequency structure of variation – Anttila (2002:211) refers to these two aspects of variation as the locus and degree of variation, and claims that an adequate theory of variation should be able to account for both. However, most models of variation developed in theoretical phonology are exclusively grammatical models – i.e. there is no place in these models where non-grammatical factors can also impact variation. (Though, see van Oostendorp 1997 for a suggestion about how non-grammatical factors could be incorporated.) In this sense, these new developments in theoretical phonology still lagged behind variationist sociolinguistics in accounting for only part of the phenomenon of phonological variation.

In this paper, I argue that the earlier research in both theoretical phonology and variationist sociolinguistics was indeed correct, and that both grammatical and non-grammatical factors influence phonological variation. Counter to most of the research in theoretical phonology, and more in line with the

---

[3] The variationist tradition did not necessarily reject all of the presuppositions of Lexical Phonology. Guy's account of the morphological influence on t/d-deletion (Guy 1991), for instance, depends crucially on the division between lexical and post-lexical rules. At the same time, however, his account also depends on the assumption that rules at each of the derivational stages are variable. The main departure from more traditional theoretical phonology was thus in rejecting the notion that *only* post-lexical rules can be optional.

variationist sociolinguistic tradition, I also argue that an adequate model of phonological variation should account for both the influence of grammatical and non-grammatical factors. However, the specific model that I propose in this paper diverges from the classic Labovian variationist model in several respects. First, it is developed in a constraint-based framework, in line with recent theoretical phonology models of variation. Secondly, grammatical and non-grammatical factors are formally separated in the model, and grammatical factors are given precedence over non-grammatical factors. Grammar alone determines whether variation is possible in a specific context, and grammatical and non-grammatical factors jointly determine the frequency structure of variation.

In the rest of this introductory section, I briefly review some of the evidence that both grammatical and non-grammatical factors influence phonological variation. The next two sections of the paper are then dedicated to the development of the model proposed here. In §2, I first show how the contribution of grammar can be accounted for in (noisy) Harmonic Grammar (Smolensky & Legendre 2006; Boersma & Pater 2008; Pater *to appear*). Section §3 then shows how the influence of non-grammatical factors can be incorporated via scaling the weights of faithfulness constraints. Section §4 briefly discusses some alternatives to the model proposed here, and §5 considers some of the implications of the model, as well as some outstanding issues.

1.1. Grammatical factors

Word-final t/d variably deletes from consonant clusters in English, so that a word like *west* can be grammatically pronounced as either [wɛst] or [wɛs]. This deletion process is probably the most extensively studied variable phonological process. It has been described in detail for countless dialects of English (see Coetzee 2004 for a recent review), and even for languages other than English (see Goeman 1999 and Goeman & van Reenen 1995 on Dutch). Since this process has been studied so extensively, the factors (both grammatical and non-grammatical) that influence its application are reasonably well understood, and for this reason, I will use t/d-deletion as a case study in the rest of this paper.

Anttila (1997:44) motivates the claim that phonological grammar should account for variation as follows: "… if variation preferences are based on phonological variables, then it seems reasonable to expect phonology to make sense of them". This insight is, of course, not new to Anttila but rather echoes what has been said in the variationist tradition for decades. The same factors that condition application of "ordinary" categorical phonological rules also condition the application of variable phonological rules. This observation is also true with regard to t/d-deletion. In a 1989-paper, Labov summarizes the literature on the grammatical factors that influence t/d-deletion, and Coetzee (2004:Chapter 5) updates this with a review of some of the more recent literature. I discuss only a few of the grammatical influences on t/d-deletion here to illustrate that this process is indeed influenced by ordinary grammatical factors.

The featural makeup of the segment that precedes the t/d influences the likelihood that deletion will apply. Nearly every study of t/d-deletion reports on how the identity of the preceding consonant influences the rate of t/d-deletion. Although dialects differ in the details, the general trend is that t/d is more likely to delete after consonants that are more similar to t/d. This has been illustrated most clearly by Guy and Boberg (1997), whose data are summarized in Table 1.[4] They interpret this as an Obligatory Contour Principle (OCP) type effect – adjacent identical features are avoided. The OCP on segmental features is, of course, a well established part of phonological grammar that conditions many ordinary categorical processes (McCarthy 1986).

Table 1. Influence of the preceding consonant on t/d-deletion (Guy & Boberg 1997)

Preceding consonant	Example	Shared features	Deletion rate (%)
[s, ʃ, z, ʒ]	*mist*	[+cor, -son]	49
[n]	*hand*	[+cor, -cont]	46
[k, g, p, b]	*pact*	[-son, -cont]	37
[l]	*field*	[+cor]	32
[f, v]	*sift*	[-son]	29
[m , ŋ]	*seemed*	[-cont]	11

Application of t/d-deletion is also influenced by what follows the t/d. In every dialect of English in which t/d-deletion has been studied, deletion is more likely to apply before consonant-initial words (*west bank*) than before vowel-initial words (*west end*) or a major phrase boundary (as in sentence final position). Table 2 gives a small, but representative, sample of the data from the literature on this aspect of t/d-deletion. The phonological context that follows some word often functions as part of the context for the application of ordinary categorical phonological rules. In French, for example, liaison that is usually blocked before a pause (*petit* [pəti] 'small') or a consonant-initial word (*petit chien* [pətiʃjɛ̃] 'small dog'), is compulsory before vowel-initial words in many contexts (*petit animal* [pətit̲animal] (Fagyal *et al.* 2006:64-69). Similarly, in some dialects of Arabic, a vowel is obligatorily inserted between a consonant-final and consonant-initial word (/kataba l-maliku / → [katabal maliku] 'the king wrote' vs. /katabat l-malikatu/ → [katabati̲l malikatu] 'the queen wrote') (Broselow 1980; Coetzee 1998).

[4] Coetzee (2004:290-293) shows that this is for the most part consistent with the data reported in other studies.

Table 2. Percent deletion in different contexts[5]

Relative deletion rate		Pre-V *west end*	Pre-Pause *west*	Pre-C *west side*
Pre-C > Pre-Pause > Pre-V	AAVE (Washington, DC)	29	73	76
	Jamaican English	63	71	85
	New York City English	66	83	100
	Tejano English	25	46	62
	Trinidadian English	21	31	81
Pre-C > Pre-V > Pre-Pause	Philadelphia English	38	12	100
	Chicano English	45	37	62
	Columbus English[6]	39	25	49

Other typical grammatical factors that condition the application of t/d-deletion include stress placement, with more deletion from unstressed (*cubist*) than stressed syllables (*insist*), and cluster size with more deletion from tri-consonantal clusters (*tanked* [tæŋkt̲]) than from bi-consonantal clusters (*tacked* [tækt̲]) (Labov 1989; Bayley 1995). From these few examples, it is clear that t/d-deletion is conditioned by grammatical factors that also condition the application of ordinary categorical phonological rules. Whatever model of phonological variation is adopted needs to explain this link between categorical and variable processes, and in this paper I

5 The data reported in this table come from the following sources: AAVE (Fasold 1972), Chicano (Santa Ana 1991), Jamaican (Patrick 1992), New York City (Guy 1980), Tejano (Bayley 1995), Trinidad (Kang 1994), Philadelphia (Guy 1980). The Columbus data is from the Buckeye Corpus (Pitt *et al*. 2007) – see the next footnote for how this information was extracted from the corpus. The data with regard to pre-consonantal context represent a simplification. The literature actually indicates that, as with the preceding consonant, the identity of the following consonant is also relevant. In general, following consonants of higher sonority tend to inhibit deletion (i.e. more deletion from *hand book* than from *hand motion*). See Bayley (1995) for an analysis in terms of syllable contact. Similarly, if t/d can form an onset cluster with the following consonant, deletion is also inhibited (more deletion from *hand lotion* than from *hand wringing*). See Labov (1997) for an analysis in terms of resyllabification.

6 The Buckeye Corpus consists of audio recordings of over 40 hours from 40 different individuals from the Columbus area of Ohio – see Pitt *et al*. (2007) for a detailed description of the corpus. The data for Figure 1 were extracted from the corpus as follows: All words that end orthographically (i.e. in spelling) on *-Ct* or *-Cd* were selected from the corpus (i.e. words to which t/d-deletion could potentially apply). From this list were then culled words with silent letters – i.e. words that end orthographically but not phonologically on a consonant cluster (e.g. *could*, *bought*, etc.). A few additional words were culled from the list due to obvious errors in the transcription in the corpus. This left a list of 20,145 words. The phone tier of the Buckeye Corpus was then consulted to determine whether each word was transcribed with a phonetic realization of its underlying final /d/ or /t/. The context that follows each of the words was also noted, and each was classified as either V, C, or pause. The frequency of each word was extracted from CELEX (Baayen *et al*. 1995), and these frequencies were log transformed. Before transformation, a constant of 1 was added to all frequencies to avoid problems with the log of words with zero frequency in CELEX. Words were then classified into frequency bins that span 0.1 of the log-transformed frequency intervals. When an interval contained too few tokens to calculate reliable statistics, it was combined with the next interval, so that there are a few intervals that are larger than 0.1. Average deletion rate for each context was determined by averaging the deletion rate of the different frequency intervals. More information available upon request.

argue that the explanation follows from the fact that both types of processes are accounted for in exactly the same way, with exactly the same ordinary machinery of grammar.

1.2. Non-grammatical factors

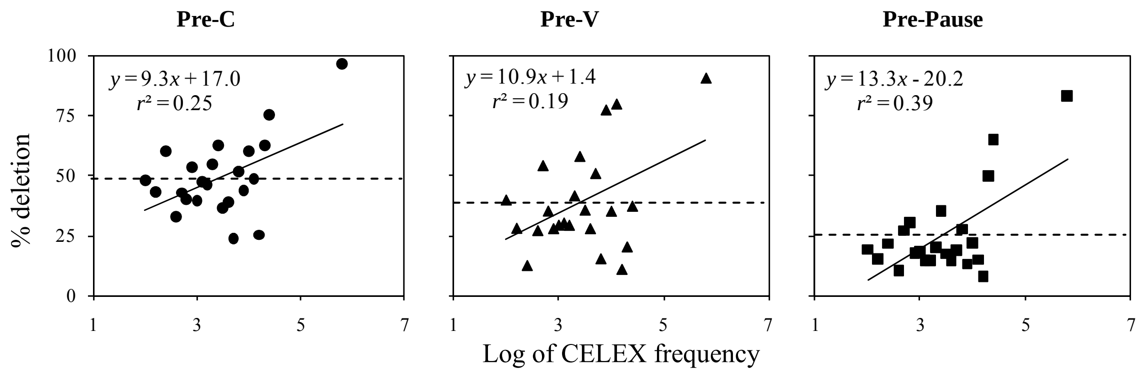
Since t/d-deletion has been studied so extensively in the variationist tradition, and since this tradition often does not make a formal distinction between grammatical and non-grammatical factors, we also have a fair amount of information on the non-grammatical factors that condition this process. As with the grammatical factors, I discuss only a few illustrative examples from the vast literature here.

Speech style or register influences the likelihood of t/d-deletion. Browman and Goldstein (1990), for instance, found little evidence of t/d-deletion applying in the reading of a word list, but they did find evidence for the process in a more casual speech style. Mitterer and Ernestus (2006) studied the analogous process in Dutch in two speech corpora. One corpus consisted of read speech (literally novels read on tape for the blind) – i.e. a rather formal reading style. The other corpus consisted of recordings of casual speech. They found evidence of deletion in both corpora, but at very different rates (8% for the read speech vs. 45% for the casual speech). Speech register is not part of grammar proper, and this is hence an example of a non-grammatical factor influencing t/d-deletion.

Frequency also interacts with t/d-deletion. In general, high frequency words (*just*) are more likely to undergo deletion than low frequency words (*bust*). Many of the variationist studies of t/d-deletion exclude high frequency words like *and*, *just* and *n't* because these words often show near categorical deletion (Patrick 1992:172; see also Bybee 2000:69-70 & 2002).

As an illustration of how frequency interacts with t/d-deletion, Figure 1 plots the deletion rate in the Pre-V, Pre-C and Pre-Pausal contexts for the English spoken in Columbus, Ohio. The *x*-axis represents the log of the CELEX (Baayen *et al.* 1995) frequency. From this figure, it is clear that deletion rate is positively correlated with frequency in all three contexts. This was confirmed by performing a linear regression for each context. The regression lines and equations, as well as the associated r^2-values, are also shown on the graphs. The slopes of all three graphs are positive and roughly equivalent. The r^2-values are also reasonably high, showing that frequency does account for a significant portion of the variation in deletion rate. The broken horizontal line represents the average deletion rate in each context, as also reported in Table 2.

Figure 1. Interaction between deletion rate and frequency in Columbus English[7]



The usage frequency of a word is obviously not part of grammar proper. The frequency of some word cannot be predicted from its phonological properties, as is shown by words like *just* and *bust* that are analogous in all relevant phonological properties but that differ drastically in their usage frequency. It is hence clear that factors other than mere grammar also influence the application of the variable process of t/d-deletion.

Any successful model of the competence that speakers have with regard to t/d-deletion needs to have the following properties: (i) It must allow phonological grammar to influence the process. (ii) But it must also allow non-grammatical factors to influence the process. (iii) And it must have an explicit theory of how the contributions from these two different sources are integrated. In the rest of this paper, I develop an Harmonic Grammar model with these properties.

2. The contribution of phonological grammar to t/d-deletion

As shown above in §1.1, there are multiple grammatical factors that influence the likelihood of t/d-deletion. I will develop an account only for the influence of the following context here – i.e. for the observation that deletion is always more likely before a consonant-initial word (*west bank*) than before a vowel-initial word (*west end*) or a phrase boundary. Relevant data are given above in Table 2. The grammatical model in which I will work, is that of (noisy) Harmonic Grammar (HG), and before I commence with developing the account, I first explain briefly what Harmonic Grammar is (§2.1). Section §2.2 then presents the account.

[7] See the previous footnote for more on how the data represented in these graphs were obtained.

2.1. (Noisy) Harmonic Grammar

Harmonic Grammar is a close relative of Optimality Theory (OT), and like OT is a grammatical model of constraint interaction rather than rule-based derivation (Legendre *et al.*2006; Pater *to appear,* Smolensky & Legendre 2006). The main difference between HG and OT is that HG uses weighted constraints rather than ranked constraints.[8] In (1) is an example of an HG and OT tableau for a language that allows no tautosyllabic clusters, and that repairs these via deletion.

(1) Optimality Theory

‘west’	DEP	*COMPLEX	MAX
west		*!	
☞ wes			*
wes.tə	*!		

Harmonic grammar

‘west’	5 DEP	1.5 *COMPLEX	1 MAX	H
west		-1		-1.5
☞ wes			-1	-1
wes.tə	-1			-5

In an HG tableau, constraint weights are indicated above the constraints in the first row. Constraint violations are marked with negative whole numbers rather than asterisks. The harmony score (H) for each candidate is then calculated by multiplying the weight of each constraint violated by the candidate by the number of violations that the candidate earns in terms of that constraint, and summing all of the values. In this tableau, for instance, [wɛst] violates DEP and MAX zero times, and *COMPLEX once. H of [wɛst] is hence calculated as follows: (*Weight*(DEP) × 0) + (*Weight*(*COMPLEX) × -1) + (*Weight*(MAX) × 0) = (5 × 0) + (1.5 × -1) + (1 × 0) = -1.5. The candidate with the highest H is selected as the optimal candidate (since all H's are negative, the highest H is always the one closest to zero).

Noisy HG (Jesney 2007; Boersma & Pater 2008) is a specific implementation of this HG model that adds a random noise component to evaluation (similar to the noise component in constraint ranking in Stochastic OT as implemented by Boersma & Hayes 2001). Every time that the grammar is used to evaluate output candidates for some input, the weight of each constraint is perturbed by a normally distributed random positive or negative number. Because of this, two constraints with weights that are sufficiently close together can lead to variation. In (2), the HG tableau from (1) is repeated, this time with the random noise added to the weight of the constraints (in parentheses next to the weight). In the first tableau, the weight of *COMPLEX is adjusted down by the addition of noise at -0.4. The weight of MAX, on the other hand, is adjusted upward by a

[8] This is more than a mere technical difference – because of this difference OT and HG differ in the kinds of languages that they predict to be possible. Specifically, “gang effects” between low weighted constraints can overpower a high weighted constraint in HG (Pater *to appear*), but in OT no combination of lower ranked constraints can ever overpower a higher ranked constraint.

positive noise value of 0.2. The effect is that violation of *COMPLEX is less serious than violation of MAX in this tableau so that the faithful candidate is selected as optimal over the deletion candidate. In the second tableau, the weight of MAX is adjusted down and that of *COMPLEX is adjusted up. In this tableau, violation of *COMPLEX is hence worse than violation of MAX, and the deletion candidate is selected as optimal.

(2) Faithful candidate optimal

‘west’	5 (.1) DEP	1.5 (-.4) *COMP	1 (.2) MAX	Ꮋ
☞ west		-1		-1.1
wes			-1	-1.2
wes.tə	-1			-5.1

Deletion candidate optimal

‘west’	5 (.2) DEP	1.5 (.4) *COMP	1 (-.2) MAX	Ꮋ
west		-1		-1.9
☞ wes			-1	-.8
wes.tə	-1			-5.2

Noisy HG has a learning algorithm analogous to the gradual learning algorithm (GLA) associated with Stochastic OT (Boersma & Hayes 2001). The HG algorithm has been implemented in *Praat* (Boersma & Weenink 2009). Like the Stochastic OT GLA, the learning algorithm associated with HG is: (i) *Error driven*. It compares the grammar’s current prediction with the learning input and adjusts the grammar only if there is a mismatch between these two. (ii) *Gradual*. The weights of the constraints are adjusted in small, incremental steps. This means that multiple cycles through the algorithm and learning data will be necessary before the final grammar is reached. See especially Boersma and Pater (2008) for a detailed discussion of the noisy HG learning algorithm.

2.2. The t/d-deletion grammar

In this section of the paper, I develop a noisy HG account for the data represented in Table 2 above. The analysis is developed here as if only the phonological/grammatical context influences the probability of deletion. In section §3, I will augment the account to allow for the influence of non-grammatical factors by scaling the weights of the faithfulness constraints.

2.2.1. The constraints

The constraints that I assume are given in (3). The two contextual faithfulness constraints are in the spirit of Steriade’s “licensing by cue” constraints – i.e. they protect segments from deletion in contexts where the cues for their perception are particularly saliently licensed (Steriade 2001, 2008).

(3) *CT — Assign one violation for every word that ends on the sequence […Ct] or […Cd].[9]

MAX — Assign one violation mark for each segment in the input that does not have a correspondent in the output (no deletion)

MAX-PRE-V — Assign one violation mark for each segment that appear in pre-vocalic context in the input, and that does not have a correspondent in the output (no deletion before a vowel)

MAX-PRE-PAUSE — Assign one violation mark for each segment that appear in pre-pausal context in the input, and that does not have a correspondent in the output (no deletion before a pause)

Steriade proposes that a segment is protected by special faithfulness constraints in contexts where its perceptual cues are more robustly licensed. The consonantal release burst can cue both place (Lahiri *et al.* 1984; Malécot 1958; Stevens & Keyser 1989; etc.) and manner information (Stevens & Keyser 1989). The formant transitions out of a consonant also carry information both about place (Celdran & Villalba 1995; Eek & Meister, 1995; Fowler 1994; Fruchter & Sussman 1997; Kewley-Port 1983; Kewley-Port *et al*. 1983; Nearey & Shammass, 1987; Stevens & Blumstein 1978; Sussman *et al*. 1991; Sussman & Shore 1996; etc.) and manner (Diehl & Walsh 1989; Walsh & Diehl 1991). To motivate the existence of the positional versions of MAX, it is therefore necessary to show that release bursts and formant transitions are more robustly licensed in pre-vocalic and pre-pausal position than in pre-consonantal position.

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

In pre-pausal position, formant transitions can obviously be realized. However, some dialects of English do strongly release final consonants (Holmes 1995), and in such dialects the release burst is licensed in pre-

[9] This constraint is a special version of the more general *COMPLEX, that applies only to a subclass of consonant clusters, and only when these clusters appear in word-final position. As it stands, the constraint is probably too specific. For instance, deletion of [p] from words like *ramp*, *wisp*, etc., and deletion of [k] from words like *whisk*, *task*, etc. is also observed. To account for these deletions, the constraint should probably be generalized so that it penalizes all […C+stop] sequences. However, the literature contains virtually no information on the deletion of [p] and [k] in these contexts, probably because there are so few […Cp] and […Ck] words in English. For this reason, I assume the more specific constraint here. However, see Coetzee (2004:252-255) for an exploration of a more general constraint.

pausal position.[10] In pre-vocalic position, both releases and transitions can be realized, but only across a word boundary. In pre-pausal position, only the release can be realized, but it is not realized across a word boundary. In pre-vocalic position, both releases and formant transitions can be realized, but only across a word boundary. Although both of these contexts licenses perception of t/d more robustly than pre-consonantal context, it is not clear whether there is a universal robustness difference between these two contexts. The robustness difference between them in some language depends on how likely final stops are to be released in this language, and on how likely releases and formant transitions are to be realized across word boundaries in this language. This is also reflected in the data reported in Table 2. Although all dialects show most deletion in pre-consonantal contexts, dialects can differ in whether pre-vocalic or pre-pausal context shows the lowest deletion rate.

2.2.2. The learning simulations

In order to develop a noisy HG account of the data in Table 2, Coetzee and Pater (*to appear*) ran learning simulations in *Praat* for each of the dialects, except for Columbus. In creating the input files for learning, they assumed that each of the three contexts (pre-consonantal, pre-vocalic, and pre-pausal) appeared 100 times. Deletion and non-deletion outputs were represented in the 100 tokens of each context proportional to the deletion rates reported in Table 2. For Chicano English, for instance, 55 of the 100 instances of pre-vocalic position was coded as having a non-deletion output and 45 as having a deletion output, etc. The decision strategy in *Praat*'s learning module was set to “Linear OT”, which is *Praat*'s implementation of noisy HG. The initial weight of all constraints was set to 100. All other settings were kept at *Praat*'s defaults. Once the grammar has been learned, *Praat*'s “to output distribution” function was used to test the predicted output of the grammar. Each of the input types was submitted 100,000 times to the grammar.[11] A grammar based on the Columbus data was learned and tested in the same way. Table 3 gives the results of the learning simulations and the testing of the grammars that have been learned. On the left, the weights that were learned by the *Praat* simulations are given. On the right, the observed deletion rates for each context, as well as the deletion rates expected given the grammar that has been learned, are given.

Let *w*(C) stand for the weight of some constraint. *w*(MAX) is very close to *w*(*CT) for six of the dialects (excluding only New York City and Philadelphia). Since the weight difference between these constraints is so small, it means that the weights actually used in evaluation (constraint weight + noise) will sometimes be higher for MAX and sometimes higher for *CT, resulting in variable deletion in pre-consonantal context.

[10] See Malécot (1958) for evidence that utterance final released consonants are perceived more accurately than unreleased consonants.
[11] The input files used by Coetzee and Pater (*to appear*) in the learning simulations reported here are included in “coetzee-pater-variation.zip”, which is available from the author, or from http://people.umass.edu/pater/coetzee-pater-variation.zip.

This is illustrated in (4) for Tejano English, where the same grammar (same basic constraint weights) selects a different output at two different evaluation occasions due to the fact that the noise is different between the two evaluations. For the New York City and Philadelphia dialects, *w*(*CT) is roughly 80 weight units higher than *w*(MAX). Since the evaluation noise was set to Praat's default of 2 during the simulation, it means that the noise added to constraint weights are normally distributed with a standard deviation of 2. The probability that a large enough positive noise value will be added to *w*(MAX) and a large enough negative noise value to *w*(*CT) for the weights of these constraints to invert is therefore vanishingly small, so that the deletion candidate is guaranteed to be selected as optimal nearly always.

Table 3. Results of learning simulations

	Constraint weights				Deletion rates			
	*CT	MAX-PRE-V	MAX-PRE-PAUSE	MAX		Pre-V	Pre-Pause	Pre-C
AAVE	101.0	3.9	-1.7[12]	99.0	Obs	29	73	76
					Exp	28	72	75
Jamaica	101.4	1.7	0.7	98.6	Obs	63	71	85
					Exp	63	71	85
NYC	141.1	80.9	78.9	58.9	Obs	66	83	100
					Exp	65	83	100
Tejano	100.4	3.0	0.8	99.6	Obs	25	46	62
					Exp	26	45	61
Trinidad	101.2	5.2	4.2	98.8	Obs	21	31	81
					Exp	20	29	79
Philadelphia	139.2	79.5	82.4	60.9	Obs	38	12	100
					Exp	38	12	100
Chicano	100.4	0.9	1.8	99.6	Obs	45	37	62
					Exp	45	38	62
Columbus	100.0	0.1	2.1	100.1	Obs	39	25	49
					Exp	38	24	49

[12] Although the "Linear OT" decision strategy in *Praat* can result in the learning of negative constraint weights, such weights are converted to zero when the grammar is used to generate outputs. Since constraint violations are given in negative integers, a negative constraint weight would result in a constraint that boosts the harmony score of candidates that violate the constraint.

(4) a. Tejano English: Faithfull output in pre-consonantal context

'west bank'	100.4 (-.1) *CT	99.6 (.8) MAX	3.0 (-.1) MAX-PRE-V	0.8 (.9) MAX-PRE-PAUSE	H
☞ west bank	-1				-100.3
wes bank		-1			-100.4

b. Tejano English: Deletion output in pre-consonantal context

'west bank'	100.4 (-.2) *CT	99.6 (.2) MAX	3.0 (-.3) MAX-PRE-V	0.8 (-.1) MAX-PRE-PAUSE	H
west bank	-1				-100.2
☞ wes bank		-1			-99.8

For all eight dialects, *w*(MAX) + *w*(MAX-PRE-V) is very close to *w*(*CT), and likewise for *w*(MAX) + *w*(MAX-PRE-PAUSE). As a result of noisy evaluation, the harmony score of the deletion candidate in pre-vocalic context is sometimes higher and sometimes lower than the faithful candidate, resulting in variable deletion in this context, and similarly for pre-pausal context. This is illustrated in (5) for pre-pausal context, again using the Tejano grammar.

(5) a. Tejano English: Faithfull output in pre-pausal context

'west'	100.4 (-.1) *CT	99.6 (.8) MAX	3.0 (1.1) MAX-PRE-V	0.8 (-.3) MAX-PRE-PAUSE	H
☞ west	-1				-100.3
wes		-1		-1	-100.9

b. Tejano English: Deletion output in pre-pausal context

'west'	100.4 (-.1) *CT	99.6 (.2) MAX	3.0 (-.3) MAX-PRE-V	0.8 (-.4) MAX-PRE-PAUSE	H
west	-1				-100.3
☞ wes		-1		-1	-100.2

Comparing the observed and the expected deletion rates for each of the dialects shows that the grammars learned in the simulation do very well at modeling the actually observed variation. It is clear that noisy HG as implemented here can account for the variation observed in these data. In addition to accounting for the data from individual dialects, the analysis developed here also expresses some universals. As shown in (6), deletion in pre-consonantal context violates a subset of the constraints violated by deletion in pre-vocalic and pre-pausal contexts. As a consequence, the harmony score of the deletion candidate in pre-consonantal position will always be higher than that of a deletion candidate in pre-pausal or pre-vocalic position. From this follows that deletion is predicted to always be most likely in pre-consonantal context.

(6) Different contexts in Tejano English

	100.4 *CT	99.6 MAX	3.0 MAX-PRE-V	0.8 MAX-PRE-PAUSE	Ⲏ
/west bank/ → [wes bank]		-1			-99.6
/west end/ → [wes end]		-1	-1		-102.6
/west/ → [wes]		-1		-1	-100.4

I conclude that noisy HG can successfully account for variable phenomena (as shown in Table 3), for categorical application of some process (as shown for pre-consonantal deletion in New York City and Philadelphia English), and for universal aspects of a variable phenomenon. However, in the account developed here, only grammar is involved. In the next section, I augment this account to allow non-grammatical factors to also influence application of the process.

3. The contribution of non-grammatical factors to t/d-deletion

As shown in §1.2 above, there are several non-grammatical factors that also influence the rate of t/d-deletion. I will focus in this section on the influence of lexical frequency as an example. See §5.1 for some exploration of how multiple non-grammatical factors can be integrated into the model.

3.1. Weight scaling

Higher frequency words are more likely than lower frequency words to undergo t/d-deletion – more deletion from high frequency *just* than low frequency *jest*. This will be the case only if the statement in (7) is true. For any input, the deletion candidate is optimal only if its harmony score (Ⲏ) is higher than that of the faithful candidate. To get more deletion in high frequency words, it is hence necessary that the probability that Ⲏ(deletion) > Ⲏ(faithful) is higher in high than low frequency words.

(7) *p*[Ⲏ(*jus_*) > Ⲏ(*just*)] > *p*[Ⲏ(*jes_*) > Ⲏ(*jest*)]

One way in which this can be achieved is by scaling the weight of faithfulness constraints down for higher frequency words, and up for lower frequency words.[13] This will entail that the faithfulness constraints are ranked lower relative to the markedness constraints for the high frequency words than for the low frequency words. In high frequency words, the markedness constraints will therefore more likely determine which candidate is optimal (higher likelihood of selecting a relatively unmarked candidate), and the opposite for low frequency words. The same effect can be achieved by scaling the weight of markedness constraints up for high frequency words, and down for low frequency words. I opt for faithfulness scaling here because there are precedents for this in the literature. In OT, for instance, van Oostendorp (1997) proposes increasingly higher ranking of faithfulness constraints in increasingly formal registers. (Though see §5.2 for possible differences between scaling markedness and faithfulness weights).

The effect of weight scaling is illustrated in (8). In these tableaux, *wt* = weight, *nz* = noise, and *sf* = scaling factor. For faithfulness constraints, there are now a weight, a noise value and a scaling factor. For markedness constraints there still are only a weight and a noise value. The constraint weights are those for Tejano English from Table 3. The two tableaux are exactly equivalent in terms of grammar (the same weights and noise values) and differ only in the scaling factor. High frequency *just* has a negative scaling factor so that the effective weights of the faithfulness constraints are lowered for *just*. The opposite holds for low frequency *jest*. With exactly the same grammatical settings, the deletion candidate is selected as optimal for high frequency *just*, and the faithful candidate for low frequency *jest*.

(8) a. Tejano English: High frequency *just*

'just'	*wt*	*nz*	*wt*	*nz*	*sf*	*wt*	*nz*	*sf*	*wt*	*nz*	*sf*	
	100.4	0.2	99.6	0.9	-1.0	3.0	-0.3	-1.0	0.8	-0.8	-1.0	
	*CT		MAX			MAX-PRE-V			MAX-PRE-PAUSE			H
just	-1											-100.6
☞ jus			-1						-1[14]			-99.5

b. Tejano English: Low frequency *jest*

'jest'	*wt*	*nz*	*wt*	*nz*	*sf*	*wt*	*nz*	*sf*	*wt*	*nz*	*sf*	
	100.4	0.2	99.6	0.9	1.0	3.0	-0.3	1.0	0.8	-0.8	1.0	
	*CT		MAX			MAX-PRE-V			MAX-PRE-PAUSE			H
☞ jest	-1											-100.6
jes			-1						-1			-102.5

[13] The idea of using weight scaling for this purpose was first suggested to me by Joe Pater (*p.c.*).
[14] When the effective weight of a constraint falls below zero, it is replaced with zero in the calculation of H to prevent a candidate from getting a harmony boost due to constraint violation.

3.2. The primacy of grammar

As shown in Table 2, pre-consonantal context has a higher deletion rate than pre-pausal and pre-vocalic context for all dialects. In §2.2.2, I have shown that this follows from the grammar of t/d-deletion. Deletion in pre-consonantal context violates a subset of the faithfulness constraints (MAX) violated by deletion in pre-pausal (MAX, MAX-PRE-PAUSE) or pre-vocalic context (MAX, MAX-PRE-V). Since weight scaling has no influence on which constraints are included in the grammar, it can also not affect the relative deletion rate differences between pre-consonantal context, and pre-vocalic or pre-pausal contexts. The grammatical factors (the constraints included in the grammar) determine the limits of what a possible language is, and the non-grammatical factors merely influence the details of how actual languages are realized within these limits.

Similarly, if in the grammar of some language the weight if MAX-PRE-V is higher than the weight of MAX-PRE-PAUSE so that deletion is more likely in pre-pausal context, the relative difference in deletion rate between these contexts cannot be affected by weight scaling. All faithfulness constraints are scaled up or down by the same scaling factor for all words of a specific frequency. The relative weight difference between MAX-PRE-V and MAX-PRE-PAUSE will therefore remain the same at all scaling factors. This is another example of the primacy of grammar – if grammar dictates more deletion in pre-pausal context via specifying a lower weight for MAX-PRE-PAUSE than MAX-PRE-V, then non-grammatical factors can never result in more deletion in pre-vocalic position.

Table 4. Deletion rates in Tejano English at different scaling factor values

Word frequency	Scaling factor	Pre-V	Pre-Pause	Pre-C
Higher	-1.0	46	64	74
	-0.5	36	55	68
	-0.1	27	47	63
Average	0	26	45	61
	0.1	24	43	60
	0.5	17	35	54
Lower	1.0	11	26	47

These two examples of grammar primacy are illustrated in Table 4. This table shows the deletion rates predicted for Tejano English under a range of different scaling factors. Deletion rates are based on the Tejano grammar from §2.2.2, and were calculated by first manually changing the weight of faithfulness constraints up or down, and then using the "to output distributions" option in *Praat*. Note that the relative deletion rate differences between the three contexts remain unchanged at all scaling factors. Also note that the deletion rate goes up as the scaling factor decreases (and lexical frequency increases), showing again how the influence of frequency is accounted for in this model.

3.3. A linking function between usage frequencies and scaling factors

The last part of the model that is missing is a linking function between frequency values and scaling factors. The details of this part of the model are still under development, so that this section of the paper is somewhat speculative. In order to develop a linking function, detailed data about deletion rates at a range of different frequency values are required. To the best of my knowledge, none of the research in the variationist tradition on t/d-deletion reports its data in such a way that this type of information can be extracted from it. The data on Columbus English represented in Figure 1, however, are of exactly the required kind, and in the rest of this section I therefore focus on Columbus English.

Using the regression equation for each of the three contexts represented in Figure 1, I determined the deletion rate predicted for a range of different frequency values. Example results are given in Table 5. I also determined the deletion rate predicted at several different scaling factors by the grammar for this dialect (as learned during the simulation discussed in §2.2.2). Example results are also given in Table 5. The next step is to align the two sets of predictions in some way. When the scaling factor is set to zero, frequency has no influence on the operation of the grammar – constraint weights and noise are then the only variables that go into the calculation of harmony scores. Based on the assumption that words of average frequency are less likely to be influenced by frequency considerations, the two sets of deletion predictions were therefore aligned so that a scaling factor of zero corresponds to the average frequency of the frequency groups in the regression analyses (a value of 3.4).

Once the two sets of values were aligned at their midpoints, the linking function in (9) was used to determine how to align the remainder of the two sets.[15] This function is a linear function so that both frequency and scaling factors change at a constant rate. In effect, the linking function in (9) does the following: (i) The scaling factor is set to zero for words with a log frequency equal to the mean log frequency of 3.4. (ii) For each increase of 0.2 above the mean log frequency, the scaling factor is decreased by 0.1. (iii) For each decrease of 0.2 below the mean log frequency, the scaling factor is increased by 0.1.

[15] For the purposes of this paper, the linking function was determined by trying several different alignments between the two sets until one was found that minimizes the differences between the two sets. In future research, ways of finding the best alignment between the two sets of value by means of an optimization algorithm should be investigated.

Table 5. Deletion rates in Columbus English, as predicted by frequency and grammar

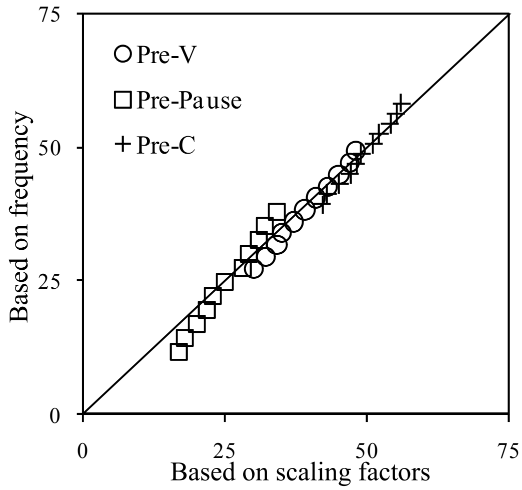
	Predicted deletion rate						
	By lg(freq)			By scaling			
lg(freq)	Pre-C	Pre-V	Pre-Pause	Pre-C	Pre-V	Pre-Pause	Scaling factor
2.4	42	30	17	39	27	12	0.5
2.6	43	32	18	41	30	14	0.4
2.8	45	34	20	43	32	17	0.3
3	47	35	22	45	34	20	0.2
3.2	48	37	23	47	36	22	0.1
3.4	49	39	25	49	38	24	0
3.6	51	41	28	51	41	27	-0.1
3.8	52	43	29	53	43	30	-0.2
4.0	54	45	31	54	45	33	-0.3
4.2	55	47	32	56	47	35	-0.4
4.4	56	48	34	58	49	38	-0.5

(9) Let *lex* be some word, f_{lex} the log frequency of *lex*, and $\mathit{diff}_{\text{lex}}$ the difference between the mean log frequency (3.4) and f_{lex} (i.e. 3.4 - f_{lex}). Let sf_{lex} be the scaling factor associated with *lex*.

$$sf_{\text{lex}} \quad = \quad (0.1)\ \frac{\mathit{diff}_{\text{lex}}}{0.2}$$

The graph in Figure 2 plots the deletion rates predicted by frequency (on the *y*-axis) and the different scaling factors (on the *x*-axis). If the two sets of predictions were perfectly matched, all data points would fall on the diagonal. The data points in Figure 2 are all tightly clustered around the diagonal, indicating a near perfect match between the two sets of predictions. That the match is this good, and that the function used to link to the two sets of predictions is a simple linear function, are encouraging. However, to conclude with more certainty that the approach proposed here for the integration of the influence of lexical frequency and grammar is successful, these results would have to be replicated for t/d-deletion in more English dialects, and also for other variable processes.

Figure 2. Deletion rates in Columbus English as predicted by frequency and weight scaling



4. Alternatives

In addition to noisy HG, there are several other models of phonological variation in the current theoretical phonology literature. In this section, I briefly consider the most important of these, and argue that noisy HG accounts better for the kind of data that is the focus of this paper. For a more detailed evaluation of these, see Coetzee and Pater (*to appear*).

4.1. Partially ordered constraints

The partially ordered constraints (POC) theory was originally proposed by Kiparsky (1993) and Reynolds (1994), and later developed in more detail by Anttila (1997, 2002). This model is developed in classic OT and thus assumed discreet rankings between constraints rather than ranking along a continuous scale. Variation arises when there are some constraints that are crucially unranked relative to each other. Every time that the grammar is used, one of the possible total rankings between the set of unranked constraints is selected at random. The frequency structure of variation also follows from this – if some variant is selected in *n* out of the *t* possible total rankings, then this variant is predicted to appear *n*/*t* of the time. There are at least three problems with this model. First, there is currently no learning algorithm associated with this model. Secondly, since constraint ranking is discreet, the range of variation that the model can predict is very limited. In this model, there are only two possible relations between any two constraints at evaluation time – either $C_1 \gg C_2$, or $C_2 \gg C_1$. The fine graded differences between the different dialects shown in Table 2 can hence not be accounted for in this model. In fact, given the same constraints as used above in §2, POC predicts that the deletion rates shown in Table 6 is the only possibility. Since deletion is observed in all three contexts, all three of the MAX-constraints must be unranked relative to each other and relative to *CT. In pre-consonantal context, there is one constraint each against deletion and retention, and deletion and retention are hence equally likely. In the other two contexts there are twice as many constraints against deletion as against retention, and retention is hence twice as likely as deletion. This is clearly not in agreement with the data from Table 2, where each dialect has different deletion rates, and where the deletion rates differ quite substantially from those predicted by POC.

Table 6. Deletion rates predicted for all dialects in the POC theory of variation

	Pre-V	Pre-Pause	Pre-C
Constraints against retention	*CT	*CT	*CT
Constraints against deletion	MAX, MAX-PRE-V	MAX, MAX-PRE-PAUSE	MAX
Expected deletion (%)	33	33	50

Thirdly, also due to the discreet ranking between constraints, there is very limited ability to allow non-grammatical factors to impact deletion rates. *CT and MAX are unranked relative to each other, allowing variable deletion in pre-consonantal position. The higher likelihood of deletion in high frequency words could be accounted for by fixing the ranking *CT ≫ MAX for high frequency words, and the lower likelihood of deletion for lower frequency words by fixing the ranking MAX ≫ *CT for such words. However, this would predict that words of average frequency show 50% deletion, high frequency words 100%, and low frequency words 0%. This is again too restrictive and not in agreement with the data.

4.2. Rank-ordered candidate set

Coetzee (2004, 2006) develops a different OT-model of variation. This model diverges from classic OT and from POC in that EVAL (the evaluative component of an OT grammar) is assumed to impose a well-formedness rank-ordering on the full candidate set rather than merely selecting the single best candidate. In this respect, it is more similar to HG where each candidate receives a harmony score so that the whole candidate set can also be ordered relative to each other. This model is dubbed the "rank-ordering model of EVAL", or ROE for short, by Coetzee. In ROE, variation arises when the language user accesses more than just the top-most candidate on the rank-ordered candidate set (more than just the optimal candidate of classic OT). The probability that some candidate will be accessed as output is proportional to its position on the rank-ordering – a candidate higher on the rank-ordering (a more well-formed candidate) is more likely to be accessed. The influence of non-grammatical factors could in principle be incorporated into this model by allowing such factors to increase or decrease the probability that candidates from lower on the rank-ordering can be accessed. However, this model has a major drawback, namely that it relies on discreet rankings like classic OT and POC. Consequently, EVAL imposes the rank-ordering $cand_1 \succ cand_2$ on two candidates, but it cannot stipulate how much better $cand_1$ is than $cand_2$. Concretely, in terms of t/d-deletion, since retention is more likely than deletion both in pre-vocalic and pre-pausal contexts in Tejano English, ROE will impose the rank-ordering *retention* ≻ *deletion* for both of these contexts. It can therefore not differentiate between these contexts, even though they do differ in deletion rates. Noisy HG, which assigns a specific numeric harmony score to every candidate, successfully overcomes this shortcoming of ROE.

4.3. Stochastic Optimality Theory

Stochastic OT, as implemented by Boersma (1997) and Boersma and Hayes (2001), differs from HG and agrees with classic OT in the use of ranked rather than weighted constraints. However, it agrees with HG in that constraint ranking is along a non-discreet, continuous scale. Because of the non-discreet ranking abilities of Stochastic OT, it is not limited to one set of deletion rates like POC (Table 6). In fact, as shown in Coetzee and Pater (*to appear*), Stochastic OT can account just as well as noisy HG for all the different dialects in Table 2. Also, like HG and different than POC, Stochastic OT has a learning algorithm affiliated with it, the Gradual Learning Algorithm (GLA). However, as shown by Pater (2008), the Stochastic OT GLA is provably non-convergent. The learning algorithm associated with HG, on the other hand, has a convergence proof (Boersma & Pater 2008). For this reason alone, the HG model of variation is to be preferred over the Stochastic OT model.

4.4. Variable rules

Labov (1969) introduced the notion of a variable rule, thereby inaugurating decades of research in the variationist sociolinguistic tradition. Variable rules are ordinary phonological rewrite rules, in the spirit of SPE (Chomsky & Halle 1968), with three augmentations: (i) the rules are marked as optional, and (ii) the structural descriptions of the rules encode how the presence of different elements in the context of the rule promotes or inhibits application of the rule. (iii) The structural descriptions of variable rules can contain both grammatical and non-grammatical factors. In fact, no formal distinction is made between grammatical and non-grammatical components of a rule's structural description. Because of the incorporation of non-grammatical factors into the structural descriptions of these rules, they can account for the influence of both grammatical and non-grammatical factors on a variable process.

Several mathematical models have been proposed over the years for relating the observed application rate of a variable rule to the presence/absence of different components of the rule's context (Cedergren & Sankoff

1974; Rousseau & Sankoff 1978; etc.). The one that has become the standard in the field, and that is implemented in the widely used software packages VarbRul and Goldvarb, performs a multivariate stepwise logistic regression over observed token counts (Paollilo 2002:177; Sankoff *et al*. 2005). In this analysis, application/non-application of the rule is taken as the dependent variable, and different factors hypothesized to influence the probability of application are taken as independent/predictor variables. Given a corpus of observed tokens to which the rule could apply, and in which each token is coded for application/non-application of the rule, as well as for the value for each of the independent variables, VarbRul/Goldvarb estimates the contribution that each independent variable makes to the probability of rule application, using a maximum likelihood algorithm. The probability that a rule will apply is expressed by the formula in (10). In this formula, p_0 represents the probability of the rule applying independently from any factors in its structural description. $p_1 \ldots p_n$ stand for factors, grammatical or non-grammatical, in the context of the rule.

(10) $$p = \frac{p_0 \times \ldots \times p_n}{\left[p_0 \times \ldots \times p_n\right] + \left[\left(1 - p_0\right) \times \ldots \times \left(1 - p_n\right)\right]}$$

Since the values of $p_0 \ldots p_n$ are determined by a best-fit algorithm, this model can model any set of input data with which it is presented extremely well. See Coetzee and Pater (*to appear*) for evidence that it does at least as well as noisy HG on the dialects in Table 2. However, also because it relies on a best-fit regression algorithm it is devoid of substantive grammatical content, and it can model any set of data with which it is presented, even if the data is grammatically unnatural. Coetzee and Pater (*to appear*) show that it can account for a language like Tejano English, except that the deletion rate in pre-consonantal and pre-vocalic contexts are flipped around. This would be an unnatural system with the lowest deletion rate in pre-consonantal position. Noisy HG, on the other hand, because it relies on grammatically substantive constraints, fails to learn a grammar that can produce a language like that, even if presented with learning data having that structure. Table 7 contains the deletion rates predicted for such a dialect by a variable rule grammar and a noisy HG grammar learned based on this kind of input. See Coetzee and Pater (*to appear*) for details about how exactly these learning simulations were conducted.

Table 7. An unnatural language with variable rules and noisy HG

	Pre-V	Pre-Pause	Pre-C
Observed deletion	62%	46%	25%
Expected deletion, based on variable rules	62%	46%	25%
Expected deletion, based on noisy HG	44%	44%	44%

As a regression model, the variable rule model is excellent at describing data on variation very accurately in a mathematically exact manner. However, since it lacks grammatical substance it is not a very effective model of natural language.

5. Concluding discussion

5.1. One or multiple scaling factors?

As shown in section §1.2, there are multiple non-grammatical factors that influence phonological variation. In addition to lexical frequency, there are, for instance, speech rate, formality register, and various aspects of social knowledge. A fully adequate model of phonological variation therefore needs to be able to incorporate the influence from grammar, and from multiple non-grammatical factors. There are several ways in which multiple non-grammatical factors can be incorporated into this model. The different non-grammatical factors could make an independent contribution so that each would have its own scaling factor. The combined influence of the non-grammatical factors will then be the sum of all the scaling factors. It is also possible that the interaction between the different non-grammatical factors are more complex – for instance, it may be that frequency is less influential at very low speech rates which is associated with careful, full pronunciations. A more complex function than mere addition would then be required for determining the combined contribution of the different non-grammatical factors. To decide between these options, more research is necessary on the way in which non-grammatical factors other than frequency impact variation, and in how different non-grammatical factors interact.

5.2. Allomorphic variation

In the model presented in this paper, only faithfulness constraints are affected by weight scaling. Candidate selection that does not depend on faithfulness constraints should therefore not be affected at all by non-grammatical factors. Kager (1996) proposes that in allomorphy, both allomorphs form part of the input and the selection between the allomorphs is made based on markedness alone – whichever allomorph results in the least marked structure is selected. Assuming this approach to allomorphy and the model of variation developed in this paper, it therefore follows that allomorphic variation should be immune to the influence of non-grammatical factors.[16] Muna (van den Berg 1989) presents an example of variable allomorphy. In Muna, pronominal possession is marked with suffixes to the noun root. The third person plural possessive suffix can be realized as either [-do] or [-^{n}do]. In roots that contain no pre-nasalized consonants, [-^{n}do] is the only

[16] Thank you to Matt Wolf for pointing out this consequence of the model to me.

allowed form of the suffix. However, in roots with a pre-nasalized consonant, free variation between the two allomorphs is allowed. Examples are given in (11). Numbers in square brackets refer to page numbers in van den Berg (1989).

(11)	[galu]	‘field’	[galu-^{n}do], *[galu-do]	‘their field’	[85]
	[ra^{n}da]	‘stomach’	[ra^{n}da-^{n}do]~[ra^{n}da-do]	‘their stomach’	[36]
	[ka^{m}bele]	‘shadow’	[ka^{m}bele-^{n}do]~[ka^{m}bele-do]	‘their shadow’	[36]

The avoidance of the [-^{n}do] form in roots containing a pre-nasalized consonant is part of a more general avoidance of words with more than one pre-nasalized consonant in Muna. Coetzee and Pater (2008) use an OCP-type constraint against multiple occurrences of the feature [pre-nasal] in a single word (*PRENAS^{2}) to account for this aspect of Muna’s phonology. In a Kager-type account of allomorphy, there is a morphological constraint stating what the preferred form of the morpheme is. In this example, [-^{n}do] is clearly the preferred form – since it is the only form allowed when *PRENAS^{2} is inert, as in a word without a pre-nasalized sound. The relevant morphological constraint would therefore be USE[-^{n}do]. These two constraints fully determine allomorph selection, and neither of them are faithfulness constraints. As such, they are immune to weight scaling so that the selection between the two variants in words with a pre-nasalized consonant should be unaffected by factors such as lexical frequency.

If it indeed is the case that variable allomorph selection is sensitive to non-grammatical factors, the model developed in this paper would need to be adjusted. It may be necessary, for instance, that weight scaling should be extended to markedness constraints also.

5.3. Final evaluation

Variation is characteristic of all levels of linguistic structure, and any adequate model of the human linguistic capacity should be able to account for variation. In the past fifteen years, significant progress has been made in this regard, and there are currently several grammatical models of variation that account very successfully for variation. However, existing models are nearly exclusively grammatical in nature while variation is impacted in non-random and predictable ways by many non-grammatical factors. A more complete model of the competence of a language user therefore needs to incorporate the influence of both grammatical and non-grammatical factors. This paper presents one proposal in which this can be achieved. Initial exploration of this model has promising results, but more research is required for a substantive evaluation of the model.

References

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

Anttila, Arto. 2002. Variation and phonological theory. In J.K. Chambers, Peter Trudgill, & Natalie Schilling-Estes, eds. *The Handbook of Language Variation and Change*. Oxford: Blackwell. p. 206-246.

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

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

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

Boersma, Paul. 1997. How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, 21:43-58.

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

Boersma, Paul & Joe Pater 2008. Convergence properties of a gradual learning algorithm for Harmonic Grammar. Ms. [ROA #970.]

Boersma, Paul & David Weenink 2009. *Praat: Doing Phonetics by Computer Version 5.1.05*. [Computer program retrieved from http://www.praat.org/.]

Broselow, Ellen. 1980. Syllable structure in two Arabic dialects. *Studies in the Linguistic Sciences*, 10:13-24.

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

Cedergren, Henrietta J. & David Sankoff. 1974. Variable rules: Performance as a statistical reflection of competence. *Language*, 50:333-355.

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

Chomsky, Noam & 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.

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

Coetzee, Andries W. *to appear*. Phonological variation and lexical frequency. In *NELS 38: Proceedings of the 38th Meeting of the North East Linguistics Society*. Amherst: GLSA.

Coetzee, Andries W. & Joe Pater. 2008. Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language and Linguistic Theory*, 26:289-337

Coetzee, Andries W. & Joe Pater. *to appear*. The place of variation in phonological theory. In John Goldsmith, Jason Riggle & Alan Yu, eds. In John Goldsmith, Jason Riggle & Alan Yu, eds. *The Handbook of Phonological Theory*. 2nd edition. Oxford: Blackwell.

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

Donnegan, Patricia J. & 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.

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

Fagyal, Zsuzsanna, Douglas Kibbee, & Fred Jenkins. *French: a Linguistic Introduction*. Cambridge: Cambridge University Press.

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

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

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

Fruchter, David & Harvey M. Sussman 1997. The perceptual relevance of locus equations. *Journal of the Acoustical Society of America*, 102:2997-3008.

Goeman, A.C.M. 1999. *T-deletie in Nederlandse Dialecten. Kwantitatieve Analyse van Structurele, Ruimtelijke en Temporele Variatie*. Ph.D. dissertation, Vrije Universiteit, Amsterdam. LOT Series 26. The Hague: Holland Academic Graphics.

Goeman, A.C.M. and P.Th. van Reenen. 1985. *Word-Final t-Deletion in Dutch Dialects. The Roles of Conceptual Prominence, Articulatory Complexity, Paradigmatic Properties, Token Frequency and Geographical Distribution*. Vrije Universiteit Working Papers in Linguistics 16. Amsterdam: Vakgroep Algemene Taalwetenschap Vrije Universiteit.

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

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

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

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

Jesney, Karen. 2008. *The Locus of Variation in Weighted Constraint Grammars*. Poster presented at the Workshop on Variation, Gradience and Frequency in Phonology. Stanford, CA: Stanford University. July 2007. [http://people.umass.edu/kjesney/Jesney2007Variation.pdf.]

Kager, René. 1996. On affix allomorphy and syllable counting. In Ursula Kleinhenz, ed. *Interfaces in Phonology*. Berlin: Akademie Verlag. p. 155-171.

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

Kang, Hyeon-Seok. 1994. Variation in past-marking and the question of the system in Trinidadian English. In *CLS 30: Papers from the 30th Regional Meeting of the Chicago Linguistic Society. Volume 2: The Parasession on Variation in Linguistic Theory*. Chicago: Chicago Linguistic Society. p. 150-164.

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

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

Kiparsky, Paul. 1982. Lexical phonology and morphology. 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.

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

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 & Leo Wetzels, eds. *Variation, Change and Phonological Theory.* Amsterdam: John Benjamins. p. 145-179.

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

Legendre, Géraldine, Antonella Sorace & Paul Smolensky. 2006. The Optimality Theory-Harmonic Grammar connection. In Paul Smolensky & Géraldine Legendre, eds. *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar, Volume 2: Linguistic and Philosophical Implications*. Cambridge, MA: MIT Press/Bradford Books. p 339-402.

Malécot, André. 1958. The role of releases in identification of released final stops. *Language*, 34:370-380.

McCarthy, John J. 1986. OCP effects: gemination and antigemination. *Linguistic Inquiry*, 17:207-263.

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

Paollilo, John C. 2002. *Analyzing Linguistic Variation: Statistical Models and Methods*. Stanford: CLSI.

Pater, Joe. *to appear*. Weighted constraints in generative linguistics. *Cognitive Science*. [ROA #982.]

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

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.

Pitt, M.A., L. Dilley, K. Johnson, S. Kiesling, W. Raymond, E. Hume & E. Fosler-Lussier. 2007. *Buckeye Corpus of Conversational Speech (2nd release)* [www.buckeyecorpus.osu.edu] Columbus, OH: Department of Psychology, Ohio State University (Distributor).

Reynolds, William. 1994. *Variation and Phonological Theory*. Ph.D. dissertation, University of Pennsylvania.

Rousseau, Pascal & David Sankoff. 1978. Advances in variable rule methodology. In David Sankoff, ed. *Linguistic Variation: Models and Methods*. New York: Academic. p. 57-69.

Sankoff, David, Sali Tagliamonte, & Eric Smith. 2005. *Goldvarb X: A variable rule application for Macintosh and Windows*. Department of Linguistics, University of Toronto. [Software package available at http://individual.utoronto.ca/tagliamonte/Goldvarb/GV_index.htm.]

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

Smolensky, Paul & Geraldine Legendre 2006. *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*. Cambridge: MIT Press.

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

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

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

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

Sussman, Harvey M., Helen A. McCaffrey, & Sandra A. Matthews. 1991. An investigation of locus equations as a source of relational invariance for stop place categorization. *Journal of the Acoustical Society of America,* 90:1309-1325.

Sussman, Harvey M. & Jadine Shore. 1996. Locus equations as phonetic descriptors of consonantal place of articulation. *Perception and Psychophysics*, 58:936-946.

van den Berg, René. 1989. *A Grammar of the Muna Language*. Dordrecht: Foris.

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

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