

The Study of Variation

Josef Fruehwald

To appear in the Handbook of the History of Phonology

14 October 2017

1 Introduction

As pointed out by Coetzee (2009), the study of what is still often called “free variation” (but not by those who study it) has followed a largely parallel history to the development of Generative Phonology following SPE. Weinreich et al. (1968) stands as the programmatic document for the subfield of Variationism, and was published the same year as SPE. Since that time, variation theory has continued to develop, and some aspects of it have become partially incorporated into mainstream phonological theory. This chapter will outline the development of variation theory, how it touches on current topics in phonology, and the direction that variation theory seems to be moving into the future.

2 The origins and early development of variationism

2.1 Attack on the idiolect: Grammar as a social object

Weinreich, Labov & Herzog’s (1968) “Empirical foundations for a theory of language change” (henceforth WLH) stands as the foundational document for modern variationist study. While Labov (1963) could be cited as the earliest modern sociolinguistic study, charting the social motivation of /ay/ raising in Martha’s Vineyard, WLH is more programmatic, and theoretically explicit. They extensively critique both the Neogrammarians (Hermann Paul specifically) and Chomsky for centering the idiolect as the primary focus and theoretical idealization of linguistic inquiry.

For Paul, speakers represented their sound systems with a bipartite kinesthetic (articulatory) representation, and a “sound image” (perceptual) representation, and the entirety of Language consisted of an associationist network between these images.¹ These representations and associationist networks are only properties of individual speakers’ minds, thus making the idiolect the only scientifically principled object of study for Paul. He treated larger scale dialects or languages as researcher artifacts, effectively averages over the idiolects of whichever speakers the researcher chose to include in the averaging.

The Chomskyian idiolect that WLH critique is a much more abstract notion than Paul’s, famously summarized in Chomsky (1965):

Linguistic theory is concerned with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest and errors (random or characteristic) in applying his knowledge of the language in actual performance.

Both the theories of Paul and Chomsky allow for variation, but not in any theoretically interesting way. For Paul, principally interested in sound systems, the bipartite articulatory and perceptual representations define an underspecified target which allows for scattered phonetic realizations surrounding it. For Chomsky, variable performance from an actual speaker encumbered by memory limitations etc. is to be expected. *Structured* variation, either within individuals or larger speech communities, gets no linguistic account.

¹Weinreich summarizes these “images” as “traces in the unconscious of physically and consciously perceived utterances.”

WLH propose a different ontology for Language which identifies it first and foremost as a social object which is only *reflected* in individual idiolects. In this approach, the only coherent object of linguistic study was the **Speech Community Grammar**, which is constituted of rules that capture the orderly heterogeneity of language, specifically **Variable Rules**. It is easy to read some of the variable rules proposed by WLH as being far-fetched in their specification. For example, variable /r/ vocalization in New York City is given the following specification:

- (1) /r/ \rightarrow g[r]/ $__$ $\left\{ \begin{array}{c} \langle K \rangle \\ \langle \# \rangle \end{array} \right\}$
- (2) g[r] = f(Style, Class, Age)

The rule in (1) is a fairly straightforward string rewrite rule with the barest added complexity of a probabilistic function $g()$ over the output. The symbols $\langle K \rangle$ and $\langle \# \rangle$ specifying the target environment are fairly uncontroversial components of any formal system in phonology, representing a following consonant and a word boundary. The rule specification becomes more unorthodox in (2), where the probabilistic function $g()$ is defined as being a function $f()$ over Style, Class and Age. That is, socioeconomic class is being included within the formal system of phonology by WLH.

At first glance, this move to include socioeconomic structures in the formal system of grammar may appear exotic to some theoreticians, but is principled on the basis of WLH's ontology of language. Neither of the rules (1) and (2) are meant to be understood as properties of any speaker's mind. These are rules in the Community Grammar, and as such they need to not only capture the linguistic distribution of variants, but also their social distribution. This point is made more clearly in Labov's 2006 notes in the second edition of the Social Stratification of English in New York City:

Language as conceived in this book is an abstract pattern, exterior to the individual. In fact, it can be argued that the individual does not exist as a linguistic entity.

2.2 Developing the theory, blurring the lines

However this non-mentalistic approach to grammar and variable rules was almost immediately blurred in Labov (1969), where variable rules are framed as more precise versions of "optional" rules from Generative Phonology (i.e. mentalistic rules). I'll illustrate the proposed formalism using the well studied TD Deletion Variable, which produces the variants *mist~mis'* and *old~ol'*. The rewrite portion of the rule is easily formulated:

$$(3) \left[\begin{array}{c} \text{Cor} \\ \text{-cont} \\ \text{+obs} \end{array} \right] \rightarrow (\emptyset)$$

Enclosing the rewrite string in parentheses indicates that rule is optional, and is associated with some probability of applying, represented by φ . Properties of the contextual conditioning affects this input probability. For example, in most English dialects the TD Deletion process only occurs when there is a preceding consonant. We'll add this to the rule specification with [+C].

$$(4) \left[\begin{array}{c} \text{Cor} \\ \text{-cont} \\ \text{+obs} \end{array} \right] \rightarrow (\emptyset) / [\text{+C}] __ \#$$

According to the rule specification, if /t, d/ preceded by [+C], the probability of the rule applying remains φ , and if it is [-C], then the probability of the rule applying is 0. That is, [\pm] valued features in the contextual conditioning defines features that potentiate the application of the rule with probability φ , and the absence or incorrect valuation of these features disallow any application of the rule.

Again, in most English dialects, if the preceding segment is a strident fricative, TD Deletion becomes more likely. We'll add this to the rule with [α strident].

$$(5) \begin{bmatrix} \text{Cor} \\ \text{-cont} \\ \text{+obs} \end{bmatrix} \rightarrow (\emptyset) / \begin{bmatrix} \text{+C} \\ \alpha \text{ strident} \end{bmatrix} \text{---}\#$$

Features valued $[\alpha, \beta, \gamma \dots]$ have the effect of boosting the probability of rule application, φ higher.

Finally, there may be some conditioning factors that increase the probability of rule application to be nearly categorical. For example, some speakers have categorical deletion of /t/ when followed by /l/ in words like *swiftly* and *softly*. This can be incorporated into the rule with [*lateral]

$$(6) \begin{bmatrix} \text{Cor} \\ \text{-cont} \\ \text{+obs} \end{bmatrix} \rightarrow (\emptyset) / \begin{bmatrix} \text{+C} \\ \alpha \text{ strident} \end{bmatrix} \text{---}[*\text{lateral}]\#$$

Features valued [*] have the effect of making the rule application obligatory.

The formulation of variable rules in Labov (1969) thus provided new machinery for the valuation of features

- (7) (a) $[\pm]$ These features potentiate variable application
 (b) $[\alpha, \beta, \dots]$ These features increase the probability of variable application.
 (c) [*] These features make rule application obligatory.

A mathematical model was also defined for translating the linguistic specification of a rule like (6) into a probability φ , which I won't address in detail here. The subject of how to capture the combinatory effects of favoring and disfavoring factors into a final output φ was the subject of considerable theory development (Cedergren and Sankoff, 1974; Kay and McDaniel, 1979; Sankoff and Labov, 1979) eventually being obviated by the use of logistic regression, a subclass of Maximum Entropy models (Goldwater and Johnson, 2003).

2.3 Discoveries of the theoretical outlook

However, the more generic result that promoting and inhibiting effects on variable processes are regularly combinatorial (regardless of the specific mathematics) is crucially non-trivial. For example, TD Deletion is more likely to happen when the following segment is a consonant than a vowel (Labov et al., 1968), and this pattern of higher pre-consonantal deletion than pre-vocalic deletion is maintained across all other morphological and phonological contexts. It is not logically necessary for this to have been so. In his critique of variable rules, Bickerton (1971) brings up the fact that Labov (1969) proposes that up to 6 features may influence variable rule application. The full crossproduct of 6 bivalent features results in 64 unique linguistic contexts, which Bickerton (1971) suggests is too many unique probabilities to track and maintain the distribution of for just two variants, especially if it necessary to maintain their appropriate ordering so that in all crossproducts the ordering of effects remain the same (e.g. pre-consonantal deletion > pre-vocalic deletion). As Cedergren and Sankoff (1974) point out, it is not necessary to track this many probabilities since the the conditioning factors on variation exhibit some degree of probabilistic independence. That is, for a variable rule with 6 influencing features, there are only 6 promoting/inhibiting weights to track, and these combine in a regular way.

However, the misunderstanding of probability in Bickerton (1971) is instructive, since in contemporary approaches to variation, tracking 64 unique probabilities is not treated as being so far fetched. For example, in in Exemplar Theoretic approaches to sound systems, the core linguistic unit is the phonetic memory trace of words or phrases (Bybee, 2002; Pierrehumbert, 2002), producing as many possible probabilities as there are utterances. In a more restrictive approach, Coetzee and Pater (2011) and Moore-Cantwell and Pater (2016) have proposed lexically indexed faithfulness constraints, but don't suggest an upper limit on how many may be included in any given grammar. In principle, this means speakers could be tracking as many unique (and independent!) probabilities as there are items in their lexicons.

The fact that the conditioning factors on variation appear to be *orderly* and *combinatorial* is not just an article of faith, nor an operating assumption of variationist research, but is one of the central *results* of variationist inquiry. Despite all that may be cognitively possible, it appears that speakers decompose variable contexts

into a number of conditionally independent factors, track promoting and inhibiting weights of these factors, and recombine them in some regular way (be that additive, multiplicative, or logistic). That is, linguistic variation appears to be influenced and constrained by many of the same structures and representations that categorical alternations are, meaning variation theory is dependent on theory development from the rest of linguistics, and variation data can also serve as evidence for theory development throughout linguistics.

Labov (2006) also set out to make strong claims about the orderly and combinatorial nature of the social conditioning of linguistic variation. For example, after finding that r-vocalization in New York City was class, gender and style stratified, Labov set out to explore the following hypothesis:

If any two sub-groups of New York City speakers are ranked on a scale of social stratification, then they will be ranked in the same order by their differential use of (r).

This could be considered another way of describing what Irvine and Gal (2000) call “fractal recursivity”, except it would seem that Labov is trying to tie the stratification of (r) to a more abstract latent variable of “stratification” that is sometimes projected onto class, sometimes onto style, and sometimes onto which department store speakers were working in.

However, this approach to social compositionality has been met with mixed success, especially with the development of intersectionality theory (Levon, 2015; Crenshaw, 1989). A clear recent example is Gates’ (2017) work on Multicultural London English [MLE] (Cheshire et al., 2011) in a London high school. In this school, there is a large gendered difference in the racial/ethnic diversity of peer group networks, with girls having more homogenous and ethnically identified groups, and the boys having more diversified groups. As such, it is nearly impossible to socially decompose MLE use as a gendered, racial, or peer group effect, as these three social dimensions appear to be co-determining each other.

2.4 Critiques of variable rules

There were a number of important critiques of the Variable Rules paradigm during its development beyond doubting the capacity of human cognition to accommodate probabilities.

2.4.1 Variable Rules are not Generativist

One important critique is due to Kay and McDaniel (1979), which focused on Labov’s (1969) theoretical framing of mentalistic variable rules as more precise versions of generativist optional rules. They rightly point out that in the original generative paradigm, a grammar is meant to define the possible and impossible strings of the language. An optional rule merely allows an additional string type into the set of possible strings of the language, and in this respect, is indistinguishable from a variable rule. For example, adding the TD Deletion rule sketched in (5) to the grammar of English has the effect of allowing both [mɪst] and [mɪs] as licit outputs for *mist* in English. The probability of the rule’s application φ , and any complexity in computing it, contributes nothing additional to this fact (but see §4 on variation in OT, and gradient acceptability below).

Variable rules are designed to capture the *frequency of alternating tokens* in language use, not the licitness of types. It has been demonstrated that the frequency of use of alternating tokens is part of speaker’s arbitrary knowledge of their language, thus must be considered a component of their linguistic competence (Roberts, 1997). For example, Smith et al. (2007) and Miller (2013) have found that individual children’s frequency of use of alternating variants ([ʌʊ]~[ɜ:] and [s]~[h]~[∅], respectively), is correlated with their specific caretakers’ frequency of use of these variants. That is, children in Chile not only acquire from their caregivers the knowledge that [lapɪs], [lapɪh] and [lapɪ] are licit output strings for the lexical item *lapiz* (pencil), but have also acquired a specific *and arbitrary* frequency with which to use these strings. The variable rules paradigm maintains the identity relationship between speakers’ competence and their grammar, but necessarily moves away from the generative paradigm of defining grammars over linguistic types.

2.4.2 Arbitrary or Universal

It is important to emphasize that it is the the *arbitrariness* of these probabilities that necessitates accounting for them in speakers' competence, much like any other aspect of language. Of the real-valued probability space between 0 and 1, learners acquire a specific and arbitrary probability for vocalizing /r/, leniting /s/, or *hoose*. We can argue that these probabilities are arbitrary because not all children acquire the same one, and we can argue that they are acquired because the probability children do acquire is influenced by their linguistic environment (their primary caretaker's usage rates) (see also Labov (1989)). However, if it could be determined that some usage rates were not arbitrary, then there would be no need for speakers to represent them in their grammars or competence at all. This should be reminiscent of the debate in phonetics regarding universal vs language specific phonetics, or whether, for example, the voicing effect on vowel duration is an intrinsic (learned) or extrinsic (universal) effect (Keating, 1985).

Within the variable rules framework, this debate has been had regarding the conditioning factors on variables from at least two different approaches. Guy and Boberg (1997) suggest that the effect of the preceding segment on TD Deletion is actually a reflection of the more general Obligatory Contour Principle (OCP). The more features preceding segments share with /t, d/, the more they promote deletion (e.g. coronal fricatives promote higher deletion than other fricatives). Under such an account, it isn't necessary for a learner to acquire specialized probabilities on a feature-by-feature basis. Rather, they just simply need to apply their independent knowledge of the OCP to the variable rule. A second approach, advocated by Temple (2009) for TD Deletion specifically, is that most of the conditioning factors can be accounted for by natural coarticulatory and fast speech rules. That is, children have little to learn about TD Deletion specifically, they simply acquire the broadly applicable articulatory strategies for their language and apply them.

2.4.3 “Different ways of saying the same thing”

One of the most challenging aspects of the theoretical construct “sociolinguistic variable” is defining it. The long standing definition is that sociolinguistic variants are “different ways of saying the same thing”, and that the more abstract linguistic object that is being realized with these variants is the variable. At first glance, it appears easy to trivially apply this definition to phonological variation where we are dealing with minimally contrastive non-meaningful units. However, the variationist approach has long been applied to variation at the morphosyntactic level (Weiner and Labov, 1983; MacKenzie, 2013; Sankoff and Thibault, 1977). Lavandera (1978) criticized this move, as it is difficult to rule out the possibility of interpretive or semantic differences between, say, an English *be*-passive construction and an equivalent *get*-passive. If there are interpretive differences between two syntactic constructions, then this moves the locus of explanation for variable use out of the grammar and into the external world. It would not be fruitful to write a variable grammar to account for why people sometimes refer to domesticated house pets as *cats* and sometimes as *dogs*, nor why people may sometimes refer to the same object as a *cup* or a *bowl* depending on what it is filled with.

However, the long standing assumption that the problem of interpretive differences is reserved for morphosyntactic variation is being fairly effectively problematized by more recent work on socioindexical theory (Ochs, 1992; Eckert, 2008). For example, Campbell-Kibler (2011) argues that the two variants [ɪ] and [ɪŋ] (e.g. *walking*~*walkin'*) are associated with distinct social meanings. That is, while the variant *walking* may be associated with the meaning “educated”, *walkin'* is not necessarily associated with the meaning “not educated”, but rather with the meaning “informal”. While in any given utterance the two variants compete for usage, their indexical associations are not interchangeable. Eckert (2008) reviews more examples. This association between social meaning and specific variants goes right back to the origins of modern variationism in Labov (1963), where it was argued that the centralizing forms of /ay/ and /aw/ ([əɪ] and [əʊ], respectively) were indexically associated with a positive orientation towards traditional lifestyles on Martha's Vineyard.

This is to say that for nearly every linguistic variable (phonological or otherwise), the use of one variant over another competing variant carries interpretive differences with it, at least in the sense of “meaning” that includes socio-pragmatic interpretations. There is not yet an obvious or clear-cut way to navigate around this issue.

3 Elaborating the model with interaction

The variable rules paradigm was dramatically advanced and its potential explanatory power expanded with a number of papers on TD Deletion in the early 1990s which focused on how variable phonological rules could interact with each other and with other variable processes.

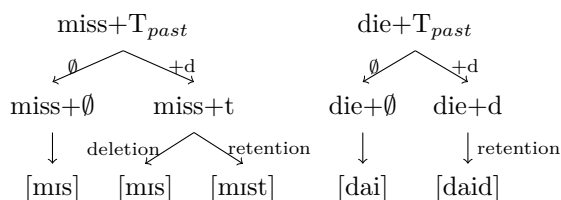
3.1 Interaction with other variable processes

Guy and Boyd (1990) and Patrick (1991) explored how phonological TD Deletion could interact with variable morphological marking to produce ambiguous TD Absence. For many English varieties, there is a consistent morphological effect on the process of TD Deletion whereby monomorphemes like *mist* undergo deletion at a higher rate than polymorphemic forms like *missed*, while semiweak past tense verbs, like *kept* undergo deletion at an intermediate rate. This is illustrated in (8).

	morphological class	example	deletion rate
(8)	monomorpheme	<i>mist</i>	highest
	semiweak	<i>kep+t, lef+t</i>	intermediate
	regular past	<i>miss#ed</i>	lowest

Patrick (1991) found that in Jamaican Creole, the regular past tense had the *highest* rate of TD Absence. While monomorphemes like *mist* had a TD Absence rate of 71%, the regular past had a TD Absence rate of 79%, reversing the ordering of most other English varieties. However, Jamaican Creole was not exceptionally different from other varieties with respect to phonological conditioning (following vowels favored retention, preceding sibilants favored absence, etc.). One distinctive property of Jamaican Creole, however, is its variable past tense marking. For example, the past tense of *give* is variably *give* (approximately 30% of the time (Patrick, 1991, Table 5)), and the past tense of *die*, which is outside the envelope of variation for TD Deletion since it ends in a vowel, is variably *die* (approximately 50% of the time). What Patrick (1991) proposed was that TD Absence in Jamaican creole was due to a mixture of morphological absence, and phonological TD Deletion, schematized in (9). The first branching node represents variable morphological marking, which affects both *miss+T_{past}* and *die+T_{past}*, which produces both marked and unmarked forms. The marked forms of *miss+T_{past}* are then variably subject to TD Deletion because of the resulting [st] cluster, which is represented by the next branching node. *Die+T_{past}*, however, is not subject to TD Deletion because its vowel final stem does not create a consonant cluster when marked, which is why the marked form is not branching in (9)

(9) Past Tense TD Absence in Jamaican Creole

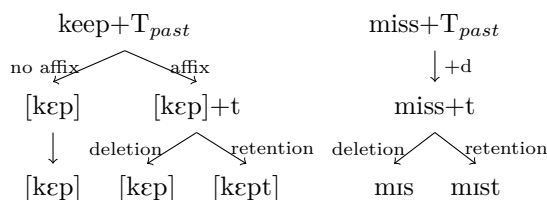


Patrick (1991) was able to use the rate of past tense marking on verbs like *die* (i.e. the probability of the first branching node) to estimate the rate of phonological deletion for the regular past tense at 60%, bringing it in line with other English varieties.

Guy and Boyd (1990) approached a very similar phenomena, whereby young children appeared to have a much higher rate of TD Absence in the semiweak verbs like *kept* and *left* than adults did. This is a pattern which has been replicated in every study to compare adults and children’s rate of TD Deletion for these verb classes (Labov, 1989; Roberts, 1997; Smith et al., 2009). Guy and Boyd (1990) proposed that this could be due to the youngest children hypothesizing that the past tense of these verbs is formed with only a stem change (e.g. *kep* and *lef*), without an additional /+t/ affix. That is, the mixture of morphological absence

and phonological deletion leads to a higher rate of TD Absence for semiweak verbs like *kept* compared to regular past tense verbs like *missed* (illustrated in (10)).

(10) Semiweak vs Regular Past tense marking (Guy and Boyd, 1990)



3.1.1 Introducing Ambiguity

With multiple potential sources of surface variants, these analyses introduced structural ambiguity to variation research. In the regular past tense in Jamaican Creole, or semiweak past tense in other varieties, if a single token is observed without a final [t, d], it is ambiguous as to whether it was morphologically absent, or phonologically deleted. In this sense, probabilistic linguistic variation is just like any other domain of linguistic inquiry: what you see is not what you get.

3.2 Cyclic interaction

Another important contribution to variation theory is Guy’s (1991) proposal that variable processes may apply multiple times within different cyclic domains, an idea incorporated from Lexical Phonology. The basic idea in Guy (1991) is that TD Deletion can only occur in word final consonant clusters, and that it cannot apply if there is a morphological bracket between the preceding consonant and the /t, d/. Thus, the morphological effect described above could be redescribed in terms of the cyclic application of a variable rule applying to strings that differ in their bracketing. This is illustrated in (11), where TD Deletion is blocked in the shaded cells by an intervening bracket.

(11) TD Deletion in Lexical Phonology

level	monomorphemes	semiweak	regular past
	<i>mist</i>	<i>kept</i>	<i>missed</i>
level 1	[[mist]]	[[kɛp]t]	[[mis]]d
level 2	[mist]	[kɛpt]	[mis]d
postlexical	mist	kept	mist

Under this analysis, TD Deletion can apply to *mist* within three cyclic domains, *kept* in two, and *missed* in just one. Guy (1991) made the strong, but only possible at the time, assumption that the probability of TD Deletion (p_{del}) applying was the same at every level, which would result in an exponential relationship in TD Retention rates.

(12) Probability of TD Retention at Each Level

level	monomorphemes	semiweak	regular past
	<i>mist</i>	<i>kept</i>	<i>missed</i>
level 1	$(1 - p_{del})$		
level 2	$(1 - p_{del})$	$(1 - p_{del})$	
postlexical	$(1 - p_{del})$	$(1 - p_{del})$	$(1 - p_{del})$
surface retention	$(1 - p_{del})^3$	$(1 - p_{del})^2$	$(1 - p_{del})$

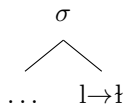
This “exponential” relationship initially appeared to be supported by the data (e.g. Santa Ana (1992)), but has not held up quantitatively since then (Smith et al., 2009; Fruehwald, 2012; Tagliamonte and Temple, 2005). However, the broader conclusion that the more cyclic domains a variable phonological process can

apply within, the higher the probability of its surface application has held up. Most notably, it has become a central theoretical plank in the Lifecycle of Phonological Change (Bermúdez-Otero, 2007, 2015), which depends on both the structural ambiguity involved in combining variable rules and cyclic domains, and the higher probability of rule application the more cyclic domains a rule can apply within.

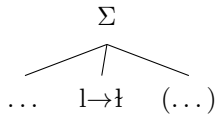
The Lifecycle of Phonological change is fundamentally a model of the (diachronic development of phonological processes, with some clear synchronic predictions for observed quantitative patterns of variation (sometimes called an ‘amphichronic’ program (Kiparsky, 2006)). One such prediction is that rules that apply in broader cyclic domains (e.g. “postlexical” or “phrase level”) should have a higher probability of application than those that apply in narrower domains (e.g. “level 1” or “stem level”). Despite the differentiation of probabilities across domains, it still holds that the more domains a rule can apply within, the higher the probability that it will have applied at the surface.

Turton (2012) is a good illustration of this more nuanced approach to cyclic interaction of variable rules. In her reanalysis of gradient well-formedness judgments from Hayes (2000), she proposes that there are the following two processes for distributing [l] and [ɫ] in English.

(13) Coda-based darkening



(14) Foot-based darkening



These two processes could have two distinct rates of application, in addition to variably applying at the stem, word, and phrase level. A broad strokes prediction is that the more processes an item is potentially subject to, the higher the rate of l-darkening, which is exactly what Turton (2012) found. For example, *bell* is potentially subject to six l-darkening processes (coda and foot based darkening in three cyclic domains), while *mail it* is only subject to 5 (all the same as *bell*, except for phrase level coda-based darkening), and in fact, *bell* has a higher rate of l-darkening than *mail it*.² A more detailed quantitative prediction is that the rates of application for each rule will be higher at the phrase level than the word level, and higher at the word level than the stem level. Turton (2012) could only estimate these different cross-cyclic rates of application for the coda-based darkening, but the prediction held up.

3.3 The contribution of “complexity”

At first glance, it may appear that incorporating models of variable morphological marking and cyclic interaction has complexified variation theory. At the very least, it brings variation theory face to face with thorny issues in morphophonology, such as exponence. It also means that accounting for the quantitative patterns in variation requires a more nuanced grammatical analysis than was perhaps previously appreciated, and that we can’t simply account for two surface variants with one rule.

However, the theoretical machinery under such an approach to variation is actually much simpler. As pointed out by Guy (1991), previously the morphological effect on TD Deletion had to be accounted for by stipulative diacritics (+, #). By eliminating these from the formalism and incorporating variable rule application into independently motivated cyclic domains, the theory is simplified. Moreover, the possibility of structural ambiguity in variable rule application, specifically the learner’s inability to attribute rule application to a specific cyclic domain on token-by-token basis allows for what Bermúdez-Otero (2007) calls “input restructuring”, or the postulation of a variable rule existing in a narrower cyclic domain than the previous generation had.

²That is, *bell* has a higher rate of l-darkening in the translation of gradient well formedness judgements to usage rates.

4 Incorporation into mainstream phonology

Phonological variability has more recently been incorporated into mainstream phonology through the development of various flavors of Harmonic Grammars, including Stochastic OT (Boersma and Hayes, 2001), Noisy Harmonic Grammars (Boersma and Pater, 2016) and Maximum Entropy Grammars (Hayes and Wilson, 2008). These approaches are well surveyed by Coetzee and Pater (2011).

It is worth emphasizing at the outset that these approaches to variation developed primarily within the core phonology literature are often trying to capture two fairly different phenomena with a single theoretical mechanism. The first is same concern of variationist sociolinguists: probabilistic linguistic output (e.g. Kiparsky, 1993; Coetzee and Kawahara, 2012; Anttila et al., 2008). The second is *gradient well-formedness*, where “consultant intuitions are gradient, falling somewhere between complete well-formedness and complete ill-formedness” (Hayes, 2000). In this re-imagining, a generative phonology does not simply define the possible and impossible output strings of the language, but rather assigns a well-formedness score of some kind to all possible output strings. In some way, this allows variationist theory to be more fully integrated into generative phonology, or at least resolves the critique from Kay and McDaniel (1979) that generative grammars are defined only over types. However, original variationist theory was exclusively concerned with the probabilistic linguistic output, and had essentially nothing to say about gradient well-formedness. It has not been suggested anywhere in the sociolinguistic literature that the less frequent variant is also in some sense less “grammatical” or somewhat ill-formed.

There are three broad approaches to accomodating variation into Harmonic Grammars like OT. They are i) underspecification, ii) stochastic re-ranking/re-weighting of constraints, and iii) translating constraint weights into probability distributions over outputs.

4.1 Underspecification


Underspecification is a very common approach to variation, whether it’s phonetic interpolation across phonologically underspecified segments (Keating, 1988), or underspecification of morphological features leading to variable agreement (Adger and Smith, 2010). In the context of Harmonic Grammars, underspecification has been proposed in the forms of Tied Constraints (Anttila, 2004) and Partially Ordered Constraints (Kiparsky, 1993). In the Tied Constraints model, two or more possible outputs have the same violations for all constraints, and are produced at random. That is, the grammar isn’t fully specified with respect to all possible outputs. To illustrate Partially Ordered Constraints, I’ll use the following constraints from Coetzee and Pater (2011) to account for the effect of the following context on TD Deletion.

- (15) (a) *Ct: Don’t have a [Ct] cluster.
 (b) MAX: Don’t delete anything.
 (c) MAX-FINAL: Don’t delete anything before a pause.
 (d) MAX-PREV: Don’t delete anything before a vowel.

A fully specified OT grammar would uniquely define a precedence relationship between all constraints. For example (16) defines only one possible complete ranking, which produced the input-output mappings in (17).

- (16) (a) MAX-PREV \gg Ct
 (b) Ct \gg MAX
 (c) MAX \gg MAX-FINAL

(17) (a)

	mist fell	MAX-PREV	*Ct	MAX	MAX-FINAL
i.	mist fell		!* *		
ii.	 mis∅ fell			*	

(b)

mist in	MAX-PREV	*Ct	MAX	MAX-FINAL
i. in mist in		*		
ii. mis in in	!*			

(c)

mist	MAX-PREV	*Ct	MAX	MAX-FINAL
i. mist		!*		
ii. in mis in			*	*

In the partially ordered constraints approach, one or more of the precedence relationships in (16) is undefined, allowing one or more constraint to “float” with respect to the others. For example, if *none* of precedence relationships between the constraints in (16) were defined, this would result in 24 possible grammars. Of these 24 possible grammars, 8 would produce TD Deletion in the string “mist in”. In the partially ordered constraints model, it is equiprobable for a speaker to be in any of the possible 24 grammar states, meaning TD Deletion would be predicted pre-vocally $\frac{8}{24} = 33\%$ of the time. Table (18) displays the predicted rates of TD Deletion for differently defined precedence relations, leaving the remainder underspecified.

Ranking	Mappings		
	{mist in → mis in in }	{mist. → mis in . }	{mist fell → mis in fell }
–	$\frac{8}{24} = 33\%$	$\frac{8}{24} = 33\%$	$\frac{12}{24} = 50\%$
MAX-PREV \gg *Ct	$\frac{0}{12} = 0\%$	$\frac{4}{12} = 33\%$	$\frac{6}{12} = 50\%$
*Ct \gg MAX-PREV	$\frac{6}{12} = 50\%$	$\frac{4}{12} = 33\%$	$\frac{6}{12} = 50\%$
MAX-FINAL \gg *Ct	$\frac{4}{12} = 33\%$	$\frac{0}{12} = 0\%$	$\frac{6}{12} = 50\%$
*Ct \gg MAX-FINAL	$\frac{4}{12} = 33\%$	$\frac{6}{12} = 50\%$	$\frac{6}{12} = 50\%$
*Ct \gg MAX	$\frac{0}{12} = 0\%$	$\frac{6}{12} = 0\%$	$\frac{6}{12} = 0\%$
MAX \gg *Ct	$\frac{8}{12} = 66\%$	$\frac{8}{12} = 66\%$	$\frac{12}{12} = 100\%$

The drawback to the partially ordered constraints model is that for any given set of precedence relationships, exactly one quantitative outcome is predicted (Coetzee and Pater, 2011). Returning to the example where all precedence relationships in (15) are underspecified, pre-vocalic TD Deletion is predicted 33% of the time, and *only* 33% of the time. The research discussed in §2.4.2 suggests that learners acquire specific and arbitrary probabilities of usage for variants, while in the partially ordered constraints approach, only one fixed probability is representable.

4.2 Stochastic Re-Ranking/Weighting

In order to incorporate arbitrary probabilities into a Harmonic Grammar, it is necessary to incorporate constraint weights in some fashion, which is what both Stochastic OT (Boersma and Hayes, 2001) and Noisy Harmonic Grammars (Boersma and Pater, 2016) do. The kind of weights that these approaches attach to their constraints are somewhat different, so I will illustrate with the Noisy-HG constraint weights learned for African American English from Coetzee and Pater (2011). Using the same constraint set (15), both a Noisy-HG and a Stochastic OT grammar may look like (19).

(19)

	*Ct	MAX	MAX-PREV	MAX-FINAL
mist	101	97	5.8	-1.5
i. mist	-1			
ii. mis∅		-1		-1

In a Stochastic-OT approach, the constraints are ranked according to their weights (as shown in (19)) and evaluated exactly like all other OT tableaux. However, during evaluation, some noise is added to the constraint weights, increasing some and decreasing others. This noise could be sufficient to rerank two similarly weighted constraints (like *Ct and MAX in (19)), resulting in a potentially different winning candidate.

Noisy-HG evaluates candidates differently. It is not the case that the candidate with the lowest ranked constraint violations wins. Rather, for each candidate, its violations are multiplied by the constraint ranking, and summed (resulting in a harmony score of -101 for *mist* and -95.5 for *mis∅* in (19)). The candidate with the largest harmony score wins. Just like Stochastic-OT, at evaluation random noise is added to the constraint weights, which affects candidates' harmony scores.

Stochastic-OT and Noisy-HG have different formal properties (e.g. Noisy-HG has constraint cumulativity). However, for the purpose of variation theory they are roughly equivalent. The arbitrary probability with which one variant may be used versus another is captured by the constraint weights. Variable output results from speakers being in different grammar states from one moment to the next. At one moment, a speaker may have a larger weight for MAX than *Ct, and at the next vice versa, resulting in variable deletion. This model of variation is actually much more similar to what is sometimes called Competing Grammars in syntactic variation (Kroch, 1989). In Stochastic-OT, it is also harder to conflate output probabilities with gradient wellformedness, since it does not produce a score of any sort for each candidate.

4.3 Generating Probability Distributions

The Harmonic Grammar model most similar quantitatively and theoretically to the Variable Rules paradigm is the Maximum Entropy (or MaxEnt) model (Hayes and Wilson, 2008). A MaxEnt Model looks very similar to a Noisy Harmonic Grammar, except rather than the candidate with the largest harmony score winning, the harmony scores are translated directly into probability distributions, then output candidates are sampled from these distributions. There is no noisy perturbation of the constraint weights in this approach. For the exact mathematical details of how this is done, see Goldwater and Johnson (2003), Hayes and Wilson (2008) or Coetzee and Pater (2011). The harmony scores in (19) would result in a probability distribution selecting *mist* 0.4% of the time, and *mis∅* 99.6% of the time. MaxEnt models have many of the same properties as Variable Rules. For example, both formalisms utilize weights of some kind on linguistic representations (phonological features for Variable Rules and phonological constraints for MaxEnt) with a mathematical function defined to combine these weights into a resulting probability. Mathematically, logistic regressions, which have been used for variable rule analysis, are a subclass of MaxEnt models. However, an important theoretical divergence between MaxEnt models and Variable Rules is MaxEnt models strongly conflate usage frequencies and gradient well-formedness, while Variable Rules were only ever meant to capture usage frequencies.

5 Aphonological Approaches

Of course, it would be impossible to discuss the approaches to phonological variation without addressing what could be called “aphonological” approaches, like Exemplar Theory. As variation data has become more popular within theoretical phonology, exemplar theoretic approaches have grown in popularity within sociolinguistic research, so much so that a special volume of the Journal of Phonetics was dedicated to the topic (Jannedy and Hay, 2006). From the observation that listeners are sensitive to and remember phonetic details of specific speakers' voices (Goldinger, 1996) (i.e. “indexical” information in the meaning of Laver

(1968)), Exemplar Theory pursues the hypothesis that speakers' knowledge of their language's sound system includes phonetically rich memory traces of specific linguistic experiences. Proposals for the exact nature of this exemplar-based knowledge differ. Bybee (2002), for example, proposes that phonetic memory traces in an associationist network with social and semantic meanings constitute speakers' *entire* knowledge of sound systems (much like Paul's theory, discussed above). Pierrehumbert (2006) advocates for a more hybridized approach.

Part of the popularity of Exemplar Theory within sociolinguistic research is undoubtedly due to its ability to capture socioindexical information in its representation (Foulkes and Docherty, 2006). Details about speaker and the social context³ are meant to be associated with the specific memory traces the speaker produced, which could go some way towards a theory of how speakers acquire knowledge about phonetic variants' indexical association (Foulkes et al., 2010). However, a preponderance of Exemplar Theoretic research focuses on the effect of word frequency on variation and change. The exact nature of frequency effects on sound change predominate the debate between proponents and critics of Exemplar Theory (Dinkin, 2008; Cohen-Goldberg, 2015; Hay et al., 2015).

The Exemplar Theoretic approach to variation is based on the production-perception feedback loop (specific approaches outlined in Pierrehumbert (2001), Wedel (2006) and Garrett and Johnson (2013)). In these models, speakers have a target production, which is either a specific phonetic memory trace sampled with some weighting from their remembered experiences, or an average generated from those remembered experiences. The eventual percept experienced by their interlocutor is not always the same as the intended production target due to transmission errors either in production or perception. These errors are not unbiased, resulting in systematic mixture of exemplars. By hypothesis, the unit of storage (i.e. the phrase, word or lexeme) will exhibit systematic effects based on its accidental long term distribution. Using TD Deletion as an example, if the word *mist* appeared in deletion favoring contexts (due to articulatory or perceptual reasons) more than the word *missed*, Exemplar Theory predicts that the long term effect is that *mist* will have more t-less target productions than *missed*. In fact, this very analysis has been proposed to account for the morphological effect on TD Deletion (Bybee, 2002; Temple, 2009). More work is necessary to establish the expected magnitude and sign of frequency effects on variation, and whether the long term distributional differences between lexical items is sufficiently different enough to produce the observed effects in variation (Cohen-Goldberg, 2015).

6 Looking forward

Whether or not Exemplar Theory is correct with respect to the nature of speakers' knowledge of sound systems, it is the case that speakers' socioindexical knowledge and word-frequency effects must be accounted for in variation theory in *some* way. Typically, the way this has been done in variation theory has been to incorporate any discovered effect on variation into the formalism, a move greatly facilitated by the conflation of statistical regression techniques with "the grammar." This tendency can be seen all the way back in Weinreich et al. (1968) in the very rule for r-vocaliation reproduced in (2) above, which includes style, age, and class in the formal specification. A more contemporary example is Coetzee and Kawahara (2012), who propose that word frequency alters constraint weightings, with more frequent words having lower weighted faithfulness constraints.

However, a unitary approach to accounting for variation has not always been adhered to. At times, it has been proposed that there may be a separate "sociolinguistic monitor" (Labov et al., 2011) or "sociocultural selection device" (Preston, 2004). Moreover there are some factors that have been found to influence variation that could not be plausibly incorporated into a formal theory. For example, Tamminga (2014) found that the probability of a past tense verb undergoing TD Deletion (like *missed*) is affected by whether the speaker deleted /t, d/ on the previous past tense verb they produced, an effect that slowly decays the longer ago it was that the previous past tense verb was produced. In order to incorporate this persistence effect into a formal grammar, like a variable rule, the rule would also have to have a memory (to store the output of its previous use) and a timer. In another case, MacKenzie (2013) found that the number of words in an

³These accounts are often less clear about how the potential infinitude of sociocultural properties are meant to be decomposed and stored in speakers' finite memory.

NP affects the probability that an auxiliary will contract onto the NP. It is a continuous effect, with each additional word in the NP decreasing the probability of contraction. Incorporating this factor into a formal model would require a grammar that counted the number of words in a preceding NP, at least up to 5, at which point contraction is almost entirely blocked.

In cases like these, the factors influencing variable linguistic output don't appear to ever condition categorical alternations. This, along with other evidence, has led Tamminga et al. (2016) to propose decomposing the causal forces on variation into three factors:

- (19) (a) Social-stylistic factors (s-conditioning)
- (b) Internal linguistic factors (i-conditioning)
- (c) Psychological and psycholinguistic factors (p-conditioning)

This is a very different kind of decomposition than the earlier proposals covered in §3 that multiple variable processes may be responsible for producing observed variable output. The interaction between variable morphological marking and variable phonological deletion would both broadly fall under i-conditioning in the Tamminga et al. (2016) account. On the other hand, factors such as how recently a speaker used a specific variant, or how many words are in the preceding NP, fall under p-conditioning, as these are effects related to the psycholinguistic *planning* of utterances.

It is possible that psycholinguistic approaches to variation will become even more prevalent in the near future. The small and growing literature on the topic includes effects of production planning on [m ~ ɱ] variation (Wagner, 2011, 2012), TD Deletion (Tanner et al., 2015) and auxiliary contraction (Mackenzie, 2016). The upshot of accounting for some factors influencing variation in production planning is that they are *not* incorporated into the narrow grammar. That is, there is no need to incorporate timers, word counters (or perhaps even word frequencies) into phonological rules or constraints, as they have a separate and independently motivated causal source in speech planning.

This move to place some explanation for variation within speech planning is simultaneously adopting the position of Chomsky (1965) that some factors influencing variation are “grammatically irrelevant” and may be “random or characteristic”, while still placing them within the scope of linguistic inquiry. The fact that many approaches attempt to account for variation entirely within grammatical formalism could be partially due to the overreliance on grammatical notation as a descriptive device, but almost certainly is largely due to the dictum that that which is linguistic is grammatical. By incorporating production planning and other additional extra-grammatical factors into variation theory, grammatical theory is kept cleaner (some may say “minimalist”) while still pursuing an accountable variationist theory. To be clear, it does not seem likely that the totality of linguistic variation could be reduced to extra-grammatical factors. As discussed above, children appear to acquire specific and arbitrary probabilities of use for linguistic variants that must be included in their knowledge of their language. At the very least, grammatical factors appear to constrain and potentiate variation, as was observed by Labov (1969) in his variable rule formalism.

7 Glancing back

An interesting aspect of this research program into the dynamics of variation in individuals is that it has a strong focus on the mentalistic construal of grammar, and therefore the individual and their idiolect. The effect of individual differences is being investigated not only terms of variable production, but also in terms of listeners' sensitivity to the socioindexical meaning of variants (Wagner and Hesson, 2014). It would seem that despite the strength of Weinreich, Labov & Herzog's (1968) critique of the idiolect, or even Labov (2006) asserting that “the individual does not exist as a linguistic entity,” the field of variationism appears to be moving towards greater theorizing of the individual.

However, some other recent work broadly falling under the study of individual differences suggest that Labov was right to say that it is at the level of the speech community that linguistic coherence is found. In comparing allophony patterns of /ɹ/ and /l/, Mielke et al. (2016) find that /ɹ/ has complex and highly idiosyncratic

conditioning, while /l/ is much more systematically constrained. They attribute this to the fact that the articulatory allophones of /ɹ/ (bunched vs retroflex) are perceptually indistinct, leading speakers to adopt allophony strategies that are most natural for their own articulatory anatomies, while /l/ allophones (light vs dark) are perceptually distinct, leading to a community of speakers coordinating on a conventionalized allophony. That is, the coherence and systematicity of allophony patterns are the product of the *speech community*, not individuals.

Having reviewed the history of variationist study, it should be clear that this is not the only case of an old debate resurfacing with the same arguments being rediscovered and re-framed. In the 19th century, Paul theorized that speakers' knowledge of their language consisted of an associationist network between phonetic realizations and semantic meanings, very similar to contemporary Exemplar Theoretic models. In the 1970s, variationists utilized logistic regression as a model of variable grammars, which is mathematically identical to Maximum Entropy Grammars in the case where only two output candidates are being evaluated. While Kay and McDaniel (1979) insisted that variable rules were not Generative, because grammars only define licit and illicit strings of the language, contemporary Harmonic Grammars assign well-formedness scores to strings.

It is possible that the similarities between these cases are strictly superficial. The researchers engaged in these similar appearing research programs and questions may not have “seen” the world the same (in the sense of Kuhn (2012)). What should be clear, however, is that in exploring the history of a field like variationism, it is important not to develop a false sense of ever advancing scientific progress, nor to assume that the questions of today have somehow superceded the concerns of the past.

References

- Adger, David, and Jennifer Smith. 2010. Variation in agreement: A lexical feature-based approach. *Lingua* 120:1109–1134. URL <http://dx.doi.org/10.1016/j.lingua.2008.05.007>.
- Anttila, Arto. 2004. Variation and Phonological Theory. *The Handbook of Language Variation and Change* 206–243.
- Anttila, Arto, Vivienne Fong, Štefan Beňuš, and Jennifer Nycz. 2008. Variation and opacity in Singapore English consonant clusters. *Phonology* 25:181. URL http://www.journals.cambridge.org/abstract_S0952675708001462.
- Bermúdez-Otero, R. 2007. Diachronic Phonology. In *The cambridge handbook of phonology*, ed. Paul de Lacy, chapter 21, 497–517. Cambridge: Cambridge University Press.
- Bermúdez-Otero, Ricardo. 2015. Amphichronic Explanation and the Life Cycle of Phonological Processes. In *The oxford handbook of historical phonology*, ed. Patrick Honeybone and Joseph Salmons, 374–399. Oxford University Press. URL <http://oxfordhandbooks.com/view/10.1093/oxfordhb/9780199232819.001.0001/oxfordhb-9780199232819-e-014>.
- Bickerton, Derek. 1971. Inherent Variability and Variable Rules. *Foundations of Language* 7:457–492.
- Boersma, Paul, and Bruce Hayes. 2001. Empirical Tests of the Gradual Learning Algorithm. *Linguistic inquiry* 32:45–86.
- Boersma, Paul, and Joe Pater. 2016. Convergence properties of a gradual learning algorithm for Harmonic Grammar. In *Harmonic grammar and harmonic serialism*, ed. John J McCarthy and Joe Pater, 389–434. Sheffield: Equinox.
- Bybee, Joan. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14:261–290.
- Campbell-Kibler, Kathryn. 2011. The sociolinguistic variant as a carrier of social meaning. *Language Variation and Change* 22:423–441. URL http://www.journals.cambridge.org/abstract_S0954394510000177.

- Cedergren, Henrietta J, and David Sankoff. 1974. Variable Rules : Performance as a Statistical Reflection of Competence. *Language* 50:333–355.
- Cheshire, Jenny, Paul Kerswill, Sue Fox, and Eivind Torgersen. 2011. *Contact, the feature pool and the speech community: The emergence of Multicultural London English*, volume 15.
- Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Coetzee, Andries W. 2009. An integrated grammatical/non-grammatical model of phonological variation. In *Current issues in linguistic interfaces*, ed. Young-Se Kang, Jong-Yurl Yoon, Hyunkyung Yoo, Sze-Wing Tang, Yong-Soon Kang, Youngjun Jang, Chul Kim, Kyung-Ae Kim, and Hye-Kyung Kang, volume 2, 267–294. Seoul: Hankookmunhwasa.
- Coetzee, Andries W., and Shigeto Kawahara. 2012. Frequency biases in phonological variation. *Natural Language & Linguistic Theory* URL <http://www.springerlink.com/index/10.1007/s11049-012-9179-z>.
- Coetzee, Andries W., and Joe Pater. 2011. The place of variation in phonological theory. In *The handbook of phonological theory*, ed. Goldsmith, Riggle, and Yu, 401–434. Wiley-Blackwell, second edition.
- Cohen-Goldberg, Ariel M. 2015. Abstract and Lexically Specific Information in Sound Patterns: Evidence from /r/-sandhi in Rhotic and Non-rhotic Varieties of English. *Language and Speech* 1–27. URL <http://las.sagepub.com/content/early/2015/02/17/0023830914567168?paperoc>.
- Crenshaw, Kimberlé. 1989. Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics. *University of Chicagor Legal Forum* 140:139–167.
- Dinkin, Aaron J. 2008. The Real Effect of Word Frequency on Phonetic Variation. *University of Pennsylvania Working Papers in Linguistics* 14:97–106.
- Eckert, Penelope. 2008. Variation and the indexical field. *Journal of Sociolinguistics* 12:453–476. URL <http://doi.wiley.com/10.1111/j.1467-9841.2008.00374.x>.
- Foulkes, Paul, and Gerard Docherty. 2006. The social life of phonetics and phonology. *Journal of Phonetics* 34:409–438.
- Foulkes, Paul, James M Scobbie, and Dominic Watt. 2010. Sociophonetics. In *The handbook of phonetic sciences*, ed. William J Hardcastle, John Laver, and Fiona E Gibbon, 703–754. Blackwell.
- Fruehwald, Josef T. 2012. Redevelopment of a Morphological Class. In *Penn Working Papers in Linguistics*, ed. Josef Fruehwald, volume 18, 77–86. URL <http://repository.upenn.edu/pwpl/vol18/iss1/10/>.
- Garrett, Andrew, and Keith Johnson. 2013. Phonetic bias in sound change. In *Origins of sound change: Approaches to phonologization.*, ed. Alan C. L. Yu, 1, 51–97. Oxford University Press. URL <http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199573745.001.0001/acprof-9780199573745-chapter-3>.
- Gates, Shivonne. 2017. Diversity within the multiethnolect: Ethnicity and language variation in the context of multiculturalism. Paper presented at UKLVC11.
- Goldinger, Stephen D. 1996. Words and Voices: Episodic Traces in Spoken Word Identification and Recognition Memory. *Journal of Experimental Psychology* 22:1166–1183.
- Goldwater, Sharon, and Mark Johnson. 2003. Learning OT Constraint Rankings Using a Maximum Entropy Model. *Proceedings of the Stockholm Workshop on Variation within Optimality Theory* 111–120.
- Guy, Gregory R. 1991. Explanation in variable phonology: An exponential model of morphological constraints. *Language Variation and Change* 3:1–22. URL http://journals.cambridge.org/abstract_S0954394500000429.
- Guy, Gregory R., and Charles Boberg. 1997. Inherent variability and the obligatory contour principle. *Language Variation and Change* 9:149.

- Guy, Gregory R., and Sally Boyd. 1990. The development of a morphological class. *Language Variation and Change* 2:1–18. URL <http://journals.cambridge.org/production/action/cjoGetFulltext?fulltextid=2703712>.
- Hay, Jennifer B., Janet B. Pierrehumbert, Abby J. Walker, and Patrick LaShell. 2015. Tracking word frequency effects through 130 years of sound change. *Cognition* 139:83–91. URL <http://linkinghub.elsevier.com/retrieve/pii/S001002771500044X>.
- Hayes, Bruce. 2000. Gradient Well-Formedness in Optimality Theory. *Optimality Theory: Phonology, Syntax, and Acquisition* 88–120.
- Hayes, Bruce, and Colin Wilson. 2008. A Maximum Entropy Model of Phonotactics and Phonotactic Learning. *Linguistic Inquiry* 39:379–440. URL <http://www.mitpressjournals.org/doi/10.1162/ling.2008.39.3.379>.
- Irvine, Judith T., and Susan Gal. 2000. Language ideology and linguistic differentiation. In *Regimes of language: Ideologies, politics, and identities*, ed. Paul V. Kroskrity, 35–83. School of American Research Press.
- Jannedy, Stefanie, and Jennifer Hay. 2006. Modelling sociophonetic variation. *Journal of Phonetics* 34:405–408.
- Kay, Paul, and Chad K McDaniel. 1979. On the logic of variable rules. *Language in Society* 8:151–187.
- Keating, Patricia A. 1985. Universal phonetics and the organisation of grammars. In *Phonetic linguistics: Essays in honor of peter ladefoged*, ed. Victoria A Fromkin, 115–132. New York: Academic Press.
- Keating, Patricia A. 1988. Underspecification in phonetics. *Phonology* 275–292.
- Kiparsky, Paul. 1993. An OT Perspective on Phonological Variation. Handout from Rutgers Optimality Workshop & NAW 1994.
- Kiparsky, Paul. 2006. The Amphichronic Program vs. Evolutionary Phonology. *Theoretical Linguistics* 32:217–236. URL <http://www.degruyter.com/view/j/thli.2006.32.issue-2/tl.2006.015/tl.2006.015.xml>.
- Kroch, Anthony. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1:199–244.
- Kuhn, Thomas. 2012. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Labov, William. 1963. The Social Motivation of a Sound Change. *Word* 19:273–309.
- 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. 2006. *The social stratification of English in New York City*. New York: Cambridge University Press, second edition.
- Labov, William, Sharon Ash, Maya Ravindranath, Tracey Weldon, Maciej Baranowski, and Naomi Nagy. 2011. Properties of the sociolinguistic monitor. *Journal of Sociolinguistics* 15:431–463. URL <http://doi.wiley.com/10.1111/j.1467-9841.2011.00504.x>.
- Labov, William, Paul Cohen, Clarence Robins, and John Lewis. 1968. *The Study of the Non-Standard English of Negro and Puerto Rican Speakers in New York City: Phonological and grammatical analysis: Report on Co-operative Research Project 3288*. New York: Columbia University Press.
- Lavandera, BR. 1978. Where does the sociolinguistic variable stop? *Language in Society* 7:171–182. URL http://journals.cambridge.org/abstract_S0047404500005510.
- Laver, J D. 1968. Voice quality and indexical information. *The British journal of disorders of communication* 3:43–54. URL <http://www.ncbi.nlm.nih.gov/pubmed/5665923>.

- Levon, Erez. 2015. Integrating Intersectionality in Language, Gender, and Sexuality Research. *Language and Linguistics Compass* 97.
- MacKenzie, Laurel. 2013. Variation in English auxiliary realization: A new take on contraction. *Language Variation and Change* 25:17–41. URL http://www.journals.cambridge.org/abstract_S0954394512000257.
- Mackenzie, Laurel. 2016. Production planning effects on variable contraction in English. *University of Pennsylvania Working Papers in Linguistics* 22:1–10.
- Mielke, Jeff, Adam Baker, and Diana Archangeli. 2016. Individual-level contact limits phonological complexity: Evidence from bunched and retroflex /ɹ/. *Language* 92:101–140.
- Miller, Karen. 2013. Acquisition of variable rules: /s/-lenition in the speech of Chilean Spanish-speaking children and their caregivers. *Language Variation and Change* 25:311–340.
- Moore-Cantwell, Claire, and Joe Pater. 2016. Gradient exceptionality in maximum entropy grammar with lexically specific constraints. *Catalan Journal of Linguistics* 15:53–66.
- Ochs, Elinor. 1992. Indexing gender. In *Rethinking context*, 335–358. Cambridge University Press.
- Patrick, Peter L. 1991. Creoles at the intersection of variable processes : -t , d deletion and past-marking in the Jamaican mesolect. *Language Variation and Change* 3:171–189.
- Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In *Frequency and the emergence of linguistic structure*, ed. Joan Bybee and Paul Hopper, volume 45 of *Typological Studies in Language*, 137–157. John Benjamins.
- Pierrehumbert, Janet B. 2002. Word-specific phonetics. In *{Laboratory Phonology VII}*, 101–139. Mouton de Gruyter.
- Pierrehumbert, Janet B. 2006. The next toolkit. *Journal of Phonetics* 34:516–530. URL <http://linkinghub.elsevier.com/retrieve/pii/S009544700600043X>.
- Preston, Dennis R. 2004. Three kinds of sociolinguistics: A psycholinguistic perspective. In *Sociolinguistic variation: Critical reflections*, ed. Carmen Fought, 140—158. New York: Oxford University Press.
- Roberts, Julie. 1997. Acquisition of variable rules: a study of (-t, d) deletion in preschool children. *Journal of child language* 24:351–72. URL <http://www.ncbi.nlm.nih.gov/pubmed/9308422>.
- Sankoff, David, and William Labov. 1979. On the uses of variable rules. *Language in Society* 8:189–222.
- Sankoff, Gillian, and Pierrette Thibault. 1977. L’alternance entre les auxiliaires avoir et être en français parlé à Montréal. *Langue française* 34:81–108. URL <http://www.jstor.org/stable/41557797>http://www.persee.fr/web/revues/home/prescript/article/lfr_0023-8368_1977_num_34_1_4818.
- Santa Ana, Otto. 1992. Chicano English evidence for the exponential hypothesis: A variable rule pervades lexical phonology. *Language Variation and Change* 4:275–288.
- Smith, Jennifer, Mercedes Durham, and Liane Fortune. 2007. “Mam, my trousers is fa’in doon!”: Community, caregiver, and child in the acquisition of variation in a Scottish dialect. *Language Variation and Change* 19:63–99.
- Smith, Jennifer, Mercedes Durham, and Liane Fortune. 2009. Universal and dialect-specific pathways of acquisition: Caregivers, children, and t/d deletion. *Language Variation and Change* 21:69.
- Tagliamonte, Sali, and Rosalind Temple. 2005. New perspectives on an ol’ variable: (t,d) in British English. *Language Variation and Change* 17:281–302. URL http://www.journals.cambridge.org/abstract_S0954394505050118.
- Tamminga, Meredith. 2014. Persistence in the Production of Linguistic Variation. Phd, University of Pennsylvania.

- Tamminga, Meredith, Laurel MacKenzie, and David Embick. 2016. The dynamics of variation in individuals. *Linguistic Variation* 16:300–336. URL <http://www.jbe-platform.com/content/journals/10.1075/lv.16.2.06tam>.
- Tanner, James, Morgan Sonderegger, and Michael Wagner. 2015. Production planning and coronal stop deletion in spontaneous speech. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*.
- Temple, Rosalind A. M. 2009. (t, d): The Variable Status of a Variable Rule. In *Papers in phonetics and computational linguistics (oxford university working papers in linguistics, philology, and phonetics 12)*, ed. Ōiwi Parker Jones and Elinor Payne, 145–170. Oxford: Oxford: Faculty of Linguistics, Philology, and Phonetics, 12 edition.
- Turton, Danielle. 2012. The Darkening of English /l/: a Stochastic Stratal OT Analysis. URL <http://ling.auf.net/lingbuzz/001524>.
- Wagner, Michael. 2011. Production Planning Constraints on Allomorphy. *Canadian Acoustics* 39:160–161.
- Wagner, Suzanne Evans. 2012. Real-time evidence for age grad(ing) in late adolescence. *Language Variation and Change* 24:179–202. URL http://www.journals.cambridge.org/abstract_S0954394512000099.
- Wagner, Suzanne Evans, and Ashley Hesson. 2014. Individual Sensitivity to the Frequency of Socially Meaningful Linguistic Cues Affects Language Attitudes. *Journal of Language and Social Psychology* 33:651–666. URL <http://journals.sagepub.com/doi/10.1177/0261927X14528713>.
- Wedel, Andrew B. 2006. Exemplar models, evolution and language change. *Linguistic Review* 23:247–274.
- Weiner, E. Judith, and William Labov. 1983. Constraints on the agentless passive. *Journal of Linguistics* 19:29–58. URL http://www.journals.cambridge.org/abstract_S0022226700007441.
- Weinreich, Uriel, William Labov, and Marvin Herzog. 1968. Empirical foundations for a theory of language change. In *Directions for historical linguistics*, ed. W Lehmann and Y Malkiel. U. of Texas Press.