27 The Study of Variation

Josef Fruehwald

27.1 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 Chomsky & Halle's (1968) *The sound pattern of English (SPE)*. Weinreich, Labov, & Herzog (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.

But first, I must address a terminological issue of what I mean by 'variation'. For the purposes of this chapter, I will be using 'variation' nearly synonymously with 'the sociolinguistic variable', which at its most oversimplistic could be defined as 'different ways of saying the same thing' (see §27.2.5.3). This is not, however, the only way in which the term 'variation' is used in linguistic inquiry.

Quite often, 'variation' is used to describe the study of 'typological variation', or parametric differences between languages, related dialects, or even stable idiolectal differences between speakers. For example, Baker (2008) describes the issue of identifying the locus of syntactic variation as concerning 'the exact ways that one speaker's knowledge of language can differ from another speaker's knowledge of language'. This kind of variation will not be the topic of this chapter, and could be best described as 'variation in knowledge'.

The kind of variation which will be the focus of this chapter could be described as 'knowledge of variation', but even then more clarity is necessary, as 'variation' is sometimes used to describe allophony; for example, the fact that /t/ in American English is realized as [t] in some circumstances and [ɾ] in others (Pitt, Dilley, & Tat 2011).

Since the focus here is on the study of phonological variation, I will define variation for the purpose of this chapter in phonological terms. 'Variation' in this chapter, then, is the phenomenon whereby a person's phonology may produce different output structures at different moments for the same given input structure.

27.2 The origins and early development of variationism

When considering the intellectual history of an area of scientific inquiry, a decision must be made about how far back it is appropriate to go. Is it possible to pinpoint a moment when the problems of the field were first defined, and methodological procedure devised to address them? Or do these earlier works only superficially appear similar, obscuring the fact that these researchers 'saw' the world differently, in the sense of Kuhn

[1962] 2012, such as Democritus' Atomism philosophy in comparison to modern Atomic Theory?

In this chapter, I will be unambiguously fixing the point of origin of contemporary variationist research on Uriel Weinreich, William Labov, & Marvin I. Herzog's (1968) 'Empirical foundations for a theory of language change' (henceforth WLH). However, no research program springs forth from nothing, and WLH's arguments have a number of intellectual roots and precursor studies that are worth mentioning at this point.

27.2.1 Origins *or* Using the past to explain the present

Contemporary Variationist Sociolinguistics springs forth from debates within historical linguistics. This is made obvious by the titles of the foundational text, 'Empirical foundations for a theory of language change' and Labov's (1994, 2001, 2010) three-volume magnum opus *Principles of linguistic change*. Both WLH and Labov (1994) situate variationist inquiry within the tensions surrounding the speed and regularity of language change. The question of speed was resolved early. Prior to Labov's groundbreaking fieldwork, there was some agreement that sound change in progress was unobservable, either because it happened too gradually (Bloomfield 1933) or too instantaneously (Hockett 1958). However, Labov's empirical studies of sound change on Martha's Vineyard (Labov 1963) and New York's Lower East Side (Labov 2006, first published in 1966) decisively demonstrated that sound change in progress could be observed. Crucially, Labov's early investigations showed that the process by which languages change from stable state A to stable state B involves an intervening period of *variation* between A and B. The attempts to theorize how variation in such a period was structured and constituted in speakers' knowledge of their language became the foundation for variationist inquiry.

The question of the regularity of change, or, as Labov (1981) called it, the 'Neogrammarian controversy', is an ongoing concern up to the present day.[1] Are sound changes gradual and exceptionless, or were the dialect geographers correct that 'every word has its own history'? On this point of contention, WLH mostly took the side of the dialect geographers. This is unsurprising, since Weinreich and Herzog were both active in dialect geography research, constructing the *Language and culture atlas of Ashkenazic Jewry* based on Yiddish dialect fieldwork with diasporic Holocaust survivors (Labov 2017). 'Herzog's Corollary' that phonemic mergers spread geographically at the expense of distinctions was based on the four-way merger of /iː, i, uː, u/ to /i/ in Yiddish in Northern Poland (Herzog 1965).

Gauchat ([1905] 2005)[2] is cited by WLH and subsequent sociolinguistic literature since then as the

[1] On the Neogrammarians and the regularity of sound change see Salmons (this volume).

[2] Here, I rely on the translation by Cummins in Chambers, Cummins, & Tennant (2008). Page numbers refer to this translation.

earliest work that most closely resembles contemporary variationist research. Gauchat carefully documents the Charmey Dialect in Western Switzerland, and his paper appears to prefigure a broad range of issues in variationist research, such as perceptually driven sound change, word frequency effects, and the fact that women tend to be leaders in language change. He tracks a number of consonantal changes (e.g. [ʎ] → [y]) and vocalic changes (e.g. [ɑ°] → [ɑ̄]), and reports intra-generation and intra-speaker variation. For example, of the debucalization of [θ] among school children, he says 'when I asked them to repeat the sentence, they often came back with *vuθo*, without recalling that they had said *vuho* the first time, in the natural abandon of the unconsidered response' (p. 254). He also explores how language internal factors (as they have come to be called) influence variation; for example, prosodic factors affect the realization of the infinitive /avɛ̄/[3] as [avɛ̄ ∼ avɛʲ ∼ avi ∼ ai ∼ i].

While Gauchat's manuscript on Charmey bears a strong resemblance to a contemporary variationist study, it is still clearly distinct from the intellectual enterprise established in 1968 by Weinreich, Labov, and Herzog. Most importantly, Gauchat's characterization of speakers' knowledge of variation could be best described as 'free variation'. Of course, he does not utilize any statistical description of variant use, but the difference in understanding of intraspeaker variation seems to be deeper than just a quantitative aesthetic. He describes some variants in terms of allophonic or phonetic rules, but there is no sense that the intraspeaker variation is also rule governed. The rule governed nature of variation is the crucial pivot point in variation study that is demarcated by WLH.

27.2.2 Attack on the idiolect: Grammar as a social object

As stated above, Weinreich, Labov, & Herzog's (1968) 'Empirical foundations for a theory of language change' 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)[4] and Chomsky for centring the idiolect as the primary focus and theoretical idealization of linguistic inquiry.

For Paul, speakers represent their sound systems with a bipartite representation: a kinesthetic (articulatory) representation, and a 'sound image' (a perceptual representation), and the entirety of language consists

[3]Cognate to French *avoir*.

[4]Weinreich died before the article was published. Labov and Herzog write that the sections on Paul and Saussure are exclusively his.

of an associationist network between these images.[5] 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 in the sense focused on in this chapter, 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.

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 is 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) $/\mathrm{r}/ \rightarrow g[\mathrm{r}]/\_\_\_\left\{ \begin{array}{c} \mathrm{K} \\ \# \end{array} \right\}$

(2) $g[\mathrm{r}] = f(\text{Style, Class, Age})$

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 $\langle \mathrm{K} \rangle$ and $\langle \# \rangle$ specifying the target environment are fairly uncontroversial components of any formal system in phonology, representing a following consonant

[5] Weinreich summarizes these 'images' as 'traces in the unconscious of physically and consciously perceived utterances'.

and a word boundary, respectively. 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.

27.2.3 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 will illustrate the proposed formalism using the well studied TD Deletion variable, which produces the variants *mist∼mis'* and *old∼ol'*. It is important to note that the way features are valued in this approach have a different interpretation than the usual notation in generative rules which is described below. The rewrite portion of the rule is easily formulated as in (3), which states that a coronal obstruent stop (i.e. /t, d/) optionally deletes when it precedes a word boundary (#):

$$(3)\quad \begin{bmatrix} \text{Coronal} \\ -\text{continuant} \\ +\text{obstruent} \end{bmatrix} \rightarrow (\emptyset) / \underline{\quad\quad} \#$$

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 will indicate this in the rule by adding the specification [+consonantal] to the left context.

$$(4)\quad \begin{bmatrix} \text{Coronal} \\ -\text{continuant} \\ +\text{obstruent} \end{bmatrix} \rightarrow (\emptyset) / \left[\ +\text{consonantal} \ \right] \underline{\quad\quad} \#$$

According to the rule specification, if /t, d/ is preceded by [+consonantal], the probability of the rule applying remains φ, and if it is [−consonantal], then the probability of the rule applying is 0. That is, [±]

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 will add [α strident] to the rule, as in (5).

(5) $\begin{bmatrix} \text{Coronal} \\ -\text{continuant} \\ +\text{obstruent} \end{bmatrix} \rightarrow (\emptyset) / \begin{bmatrix} +\text{consonantal} \\ \alpha \text{ strident} \end{bmatrix} \text{\_\_\_\_} \#$

Features valued [α, β, γ ...] have the effect of boosting the probability of rule application, φ. In this case, 'α' represents a quantitative weight, boosting the probability of rule application by the specified amount, *not* feature agreement as is the more usual use of Greek letters in generative rule notation.

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 by adding [$*$lateral] (6):

(6) $\begin{bmatrix} \text{Coronal} \\ -\text{continuant} \\ +\text{obstruent} \end{bmatrix} \rightarrow (\emptyset) / \begin{bmatrix} +\text{consonantal} \\ \alpha \text{ strident} \end{bmatrix} \text{\_\_\_\_}[*\text{lateral}]\#$

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, summarized in (7).

(7) (a) [$\pm$] These features potentiate variable application

(b) [α, β, ...] 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 favouring and disfavouring factors into a final output φ was the subject of considerable theory development (Cedergren & Sankoff 1974; Kay & McDaniel 1979; Sankoff & Labov 1979) eventually being obviated by the use of logistic regression, a subclass of maximum entropy models (Goldwater & Johnson 2003).

27.2.4 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, Cohen, Robins, & Lewis 1968), and this pattern 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 remains the same (e.g. pre-consonantal deletion > pre-vocalic deletion). As Cedergren & 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 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 & Pater (2011) and Moore-Cantwell & Pater (2016) have proposed lexically indexed faithfulness constraints, but do not 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 & Gal (2000) call 'fractal recursivity', which 'involves the projection of an opposition, sailent at some level of relationship, onto some other level', 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 (Crenshaw 1989; Levon 2015). Intersectionality theory argues that people's identities cannot be neatly decomposed into constituent parts. For example, the experience of being a Black woman in America cannot be understood through a simple process of 'adding' together the experience of being Black, the experience of being a woman, and the experience of being Amercian. Rather, the conjunction of these identities creates a unique and distinctive experience. A clear recent example of the non-additive social effects on language use is Gates' (2018, 2019) work on Multicultural London English [MLE] (Cheshire, Kerswill, Fox, & Torgersen 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 homogeneous 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.

27.2.5 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.

27.2.5.1 Variable rules are not generativist

One important critique is due to Kay & 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 §27.4 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 speakers' arbitrary knowledge of their language, thus must be considered a component of their linguistic competence (Roberts 1997). For example, Smith, Durham, & Fortune (2007) and Miller (2013) have found that individual children's frequency of use of alternating variants ([ʌʊ]∼[ʉː] in words like *house* or *now* in Buckie Scots and [s]∼[h]∼[∅] in Chilean Spanish, 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 [lapis], [lapih] and [lapi] 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.

27.2.5.2 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 using [ʉː] in *house*. We can argue that these probabilities are arbitrary because not all children acquire the same ones, and we can argue that they are acquired because the probabilities children do acquire are 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).

Variationist work itself has wavered on the issue of whether the conditioning weights in variable rules are arbitrary or are perhaps natural or universal in some way. Guy & 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), which bans successive identical specifications (see Kisseberth, this volume). 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 is not 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.

27.2.5.3 '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'. This can be illustrated with the contemporary morphological doublet for the past tense of *sneak*, which appears both as *sneaked* and *snuck*. Each realization is a 'variant'. The more abstract representation which these variants are realizations of, *sneak*+$\mathrm{T}_{\mathrm{past}}$, is the 'variable'.

At first glance, one might suppose that this definition would apply most easily to phonological variation, where we are dealing with minimally contrastive units that are non-meaningful by definition. Yet the variationist approach has long been applied to variation at the morphosyntactic level as well (Sankoff & Thibault 1977; Weiner & Labov 1983; MacKenzie 2013). The *sneak* ∼ *snuck* doublet mentioned above appears to be an unproblematic morphological example. Lavandera (1978) criticized this move, however, on the grounds that 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 of referents that speakers are trying to describe. 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 assumption that the problem of interpretive differences is reserved for morphosyntactic variation has been fairly effectively problematized by more recent work on socioindexical theory (Ochs 1992; Silverstein 2003; Eckert 2008). For example, Campbell-Kibler (2011) argues that the two variants [-ɪŋ] and [-ɪn] (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 in general. For phonological variation, the best approach may be to narrowly define the 'variable' as synonymous with the probabilistic phonological process that, given the same input, generates different outputs from moment to moment. The socio-pragmatic factors quantitatively influence the output, rather than deterministically selecting a specific variant.

27.3 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.

27.3.1 Interaction with other variable processes

Guy & 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 Table 27.1.

morphological class	example	deletion rate
monomorpheme	*mist*	highest
semiweak	*kep+t*, *lef+t*	intermediate
regular past	*miss#ed*	lowest

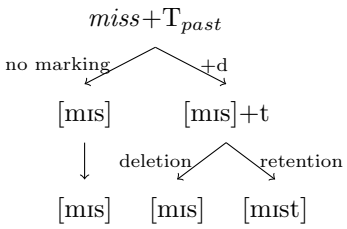
Table 27.1: Morphological Boundaries and TD Deletion Rates

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 favoured retention, preceding sibilants favoured 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

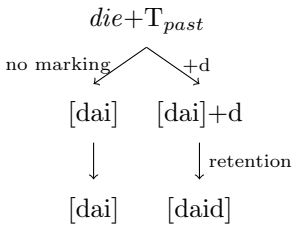
time (Patrick 1991: Table 5), and the past tense of *die*, which is outside the envelope of variation for TD Deletion since it ends in a vowel, is variably *die* (approximately 50% of the time). What Patrick (1991) proposed was that TD Absence in Jamaican Creole was due to a mixture of morphological absence and phonological TD Deletion, schematized in (8). 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$_{past}$ (8a) are then variably subject to TD Deletion because of the resulting [st] cluster, which is represented by the next branching node. *Die*+T$_{past}$ in (8b), 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 (8b).

(8) Past tense TD Absence in Jamaican Creole

(a) *miss*



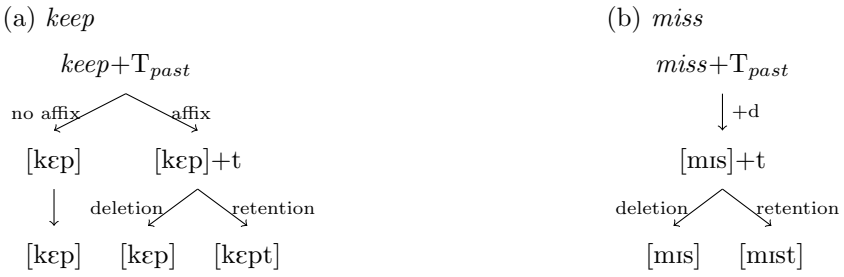
(b) *die*



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 & Boyd (1990) approached a very similar phenomenon, 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 that compared adults and children's rate of TD Deletion for these verb classes (Labov 1989; Roberts 1997; Smith, Durham, & Fortune 2009). Guy & 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* (9a) compared to regular past tense verbs like *missed* (9b).

(9) Semiweak vs regular past tense marking (Guy & Boyd 1990)



27.3.1.1 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 is 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.

27.3.2 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 (Kiparsky 1982, 1985; Kaisse & Shaw 1985; Mohanan 1986; see Scheer, this volume). 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 Table 27.2, where TD Deletion is blocked in the shaded cells by an intervening bracket.

level	monomorphemes	semiweak	regular past
	mist	*kept*	*missed*
level 1	[[mɪst]]	[[kɛp]t]	[[mɪs]]d
level 2	[mɪst]	[kɛpt]	[mɪs]d
postlexical	mɪst	kɛpt	mɪst

Table 27.2: TD Deletion in Lexical Phonology

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 assumption (though for practical reasons perhaps the only one

possible at the time) 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.

level	monomorphemes	semiweak	regular past
	mist	*kept*	*missed*
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	$(1 - p_{del})^3$	$(1 - p_{del})^2$	$(1 - p_{del})$

Table 27.3: Probability of TD Retention at Each Level

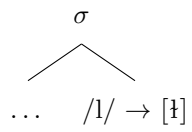
This 'exponential' relationship initially appeared to be supported by the data (e.g. Santa Ana 1992), but has not held up quantitatively since then (Smith, Durham, & Fortune 2009; Fruehwald 2012; Tagliamonte & Temple 2005). However, the broader conclusion has held up: the more cyclic domains a variable phonological process can apply within, the higher the probability of its surface application. Most notably, it has become a central theoretical plank in the life cycle of phonological processes model of 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 when a rule can apply within more cyclic domains.

The life cycle of phonological processes 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; see Kiparsky 2006). Bermúdez-Otero (2007) proposes that new phonological rules initially apply at just the phrase level phonology with a low probability of application. Then, as their probability of application increases at this broadest cyclic domain, a new version of the rule enters the next narrower cyclic domain. A prediction of this proposal is that rules that apply in broader cyclic domains should have a higher probability of application than those that apply in narrower domains since they have been present in the community's grammar for longer. 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.

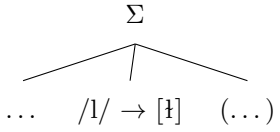
Turton (2012) is a good illustration of this more nuanced approach to cyclic interaction of variable rules. In her reanalysis of gradient well-formedness judgments from Hayes (2000), she proposes that in English there must be two separate *l*-darkening processes which target /l/ in two different prosodic contexts—within the coda (10a) and foot internally (10b)—and that both of these processes apply within three cyclic domains (stem, word, and phrase level).

(10) Two contexts for *l*-darkening

(a) *l*-darkening in coda



(b) Foot-internal *l*-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-internal darkening in three cyclic domains), while *mail it* is only subject to five (see Tables 27.4 and 27.5); in fact, *bell* has a higher rate of *l*-darkening than *mail it*.[6]

Level	*bell*	*l* in coda	*l* foot internal
stem	[bɛl]	✓	✓
word	[bɛl]	✓	✓
phrase	[bɛl]	✓	✓

Table 27.4: *l*-darkening processes that can apply to *bell*

Level	*mail in*	*l* in coda	*l* foot internal
stem	[meil]	✓	✓
word	[meil]	✓	✓
phrase	[mei.lɪn]	✗	✓

Table 27.5: *l*-darkening processes that can apply to *mail it*

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.

27.3.3 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

[6]That is, *bell* has a higher rate of *l*-darkening in the translation of gradient well formedness judgments to usage rates.

patterns in variation requires a more nuanced grammatical analysis than was perhaps previously appreciated, and that we cannot 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 (+, #).[7] 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 a 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.

27.4 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 & Hayes 2001), Noisy Harmonic Grammars (Boersma & Pater 2016) and maximum entropy grammars (Hayes & Wilson 2008). These approaches are well surveyed by Coetzee & Pater (2011); see also van Oostendorp (this volume).

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 also the concern of variationist sociolinguists: probabilistic linguistic output (e.g. Kiparsky 1993; Anttila, Fong, Beňuš, & Nycz 2008; Coetzee & Kawahara 2013). 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 & 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 accommodating variation into Harmonic Grammars like OT. They

[7]On boundary symbols as diacritics see Scheer (this volume).

are i) underspecification, ii) stochastic re-ranking/re-weighting of constraints, and iii) translating constraint weights into probability distributions over outputs.

27.4.1 Underspecification

Underspecification is a very common approach to variation, whether it is phonetic interpolation across phonologically underspecified segments (Keating 1988), or underspecification of morphological features leading to variable agreement (Adger & 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; Reynolds 1994). 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 is not fully specified with respect to all possible outputs. To illustrate partially ordered constraints, I'll use the constraints in (11) from Coetzee & Pater (2011) to account for the effect of the following context on TD Deletion.

(11) (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, (12) defines only one possible complete ranking, which produces the input-output mappings in (13).

(12) (a) Max-PreV≫Ct

(b) Ct≫Max

(c) Max≫Max-Final

(13) (a)

mist fell	Max-PreV	*Ct	Max	Max-Final
i. mist fell		!∗		
ii. ☞ misØ fell			∗	

(b)

mist in	Max-PreV	*Ct	Max	Max-Final
i. ☞ mist in		∗		
ii. misØ in	!∗		∗	

(c)

mist	Max-PreV	*Ct	Max	Max-Final
i. mist		!*		
ii. ☞ misØ			*	*

In the partially ordered constraints approach, one or more of the precedence relationships in (12) is undefined, allowing one or more constraints to 'float' with respect to the others. For example, if *none* of the precedence relationships between the constraints in (11) 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-vocalically $\frac{8}{24} = 33\%$ of the time. Table 27.6 displays the predicted rates of TD Deletion for differently defined precedence relations, leaving the remainder underspecified.

Ranking	Mappings {mist in → misØ in }	{mist. → misØ. }	{mist fell → misØ fell }
–	$\frac{8}{24} = 33\%$	$\frac{8}{24} = 33\%$	$\frac{12}{24} = 50\%$
Max-PreV≫*Ct	$\frac{0}{12} = 0\%$	$\frac{4}{12} = 33\%$	$\frac{6}{12} = 50\%$
*Ct≫Max-PreV	$\frac{6}{12} = 50\%$	$\frac{4}{12} = 33\%$	$\frac{6}{12} = 50\%$
Max-Final ≫ *Ct	$\frac{4}{12} = 33\%$	$\frac{0}{12} = 0\%$	$\frac{6}{12} = 50\%$
*Ct ≫ Max-Final	$\frac{4}{12} = 33\%$	$\frac{6}{12} = 50\%$	$\frac{6}{12} = 50\%$
*Ct ≫ Max	$\frac{8}{12} = 66\%$	$\frac{8}{12} = 66\%$	$\frac{12}{12} = 100\%$
Max ≫ *Ct	$\frac{0}{12} = 0\%$	$\frac{0}{12} = 0\%$	$\frac{0}{12} = 0\%$

Table 27.6: Predicted TD Deletion rates given just one specified precedence relation

The drawback to the partially ordered constraints model is that for any given set of precedence relationships, exactly one quantitative outcome is predicted (Coetzee & Pater 2011). Returning to the example where all precedence relationships in (11) are underspecified, pre-vocalic TD Deletion is predicted 33% of the time, and *only* 33% of the time. The research discussed in §27.2.5.2 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. Of course, the probability of selecting one set of precedence relations may not be uniformly distributed, in which case more than one fixed probability could be produced by the

system. However, if we were to opt for this approach, the mechanism of underspecification would no longer be generating the usage frequencies of variants. Rather, this would be done by the mechanism which selects a specific set of precedence relations, which has not yet been explicated in the literature.

27.4.2 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 & Hayes 2001) and Noisy Harmonic Grammars (Boersma & 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 & Pater (2011). Using the same constraint set (11), both a Noisy HG and a Stochastic OT grammar may look like (14).

(14)

	*Ct	Max	Max-PreV	Max-Final
mist	101	97	5.8	-1.5
i. mist	-1			
ii. ☞ misØ		-1		-1

In a Stochastic OT approach, the constraints are ranked according to their weights (as shown in (14)) 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 (14)), 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 necessarily 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 *misØ* in (14)).[8] The candidate with the largest (i.e. least negative) 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. However, for the purpose of variation theory they are roughly equivalent. The arbitrary probability with which one variant may be used versus

[8]Since Max-Final has a negative weight, when it is multiplied by the violation, it actually moves the harmony score to be larger.

another is captured by the constraint weights. Variable output results from speakers being in different grammar states from one moment to the next. A speaker may have a larger weight for MAX than *Ct at one moment but not at the next, 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.

27.4.3 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 & 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. This is done by exponentiating the harmony score for each candiate, then dividing by the sum of all exponentiated harmony scores (see Goldwater & Johnson (2003), Hayes & Wilson (2008), or Coetzee & Pater (2011) for more details). The harmony scores in (14) 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 that MaxEnt models strongly conflate usage frequencies and gradient well-formedness, while variable rules were only ever meant to capture usage frequencies.

27.5 Exemplar theoretic 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 (see Pierrehumbert, this volume). 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 & 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 representations (Foulkes & Docherty 2006). Details about speaker and the social context[9] 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, Scobbie, & Watt 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 dominates the debate between proponents and critics of exemplar theory (Dinkin 2008; Cohen-Goldberg 2015; Hay, Pierrehumbert, Walker, & LaShell 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 & 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 a systematically biased 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 favouring 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 different enough to produce the observed effects in variation (Cohen-Goldberg 2015).

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

27.6 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, Labov, & Herzog (1968) in the very rule for *r*-vocalization reproduced in (2) above, which includes style, age, and class in the formal specification. A more contemporary example is Coetzee & Kawahara (2013), 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 a noun phrase (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, MacKenzie, & Embick (2016) to propose decomposing the causal forces on variation into three factors:

(15) (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 §27.3 that multiple variable processes may be responsible for producing observed variable output. The interaction between vari-

able morphological marking and variable phonological deletion would both broadly fall under i-conditioning in the Tamminga, MacKenzie, & Embick (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 [ɪn] ∼ [ɪŋ] variation (Wagner 2011, 2012), TD Deletion (Tanner, Sonderegger, & Wagner 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.

27.7 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 & 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 suggests 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, Baker, & Archangeli (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 consists 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 & McDaniel (1979) insisted that variable rules are 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 [1962] 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 superseded the concerns of the past.

References

Adger, David and Jennifer Smith. 2010. Variation in agreement: A lexical feature-based approach. *Lingua* 120(5): 1109–1134. doi:10.1016/j.lingua.2008.05.007. URL http://dx.doi.org/10.1016/j.lingua.2008.05.007.

Anttila, Arto. 2004. Variation and phonological theory. In J. K. Chambers, Peter Trudgill, and Natalie Schilling-Estes (eds.), *The handbook of language variation and change*, 206–243. Oxford: Blackwell. Wiley Online Library. doi:10.1002/9780470756591.ch8.

Anttila, Arto, Vivienne Fong, Štefan Beňuš, and Jennifer Nycz. 2008. Variation and opacity in Singapore English consonant clusters. *Phonology* 25(02): 181. doi:10.1017/S0952675708001462. URL http://www.journals.cambridge.org/abstract{_}S0952675708001462.

Baker, Mark C. 2008. The macroparameter in a microparametric world. In Theresa Biberauer (ed.), *The limits of syntactic variation*, 351–74. Benjamins.

Bermúdez-Otero, Ricardo. 2007. Diachronic phonology. In Paul de Lacy (ed.), *The Cambridge handbook of phonology*, chap. 21, 497–517. Cambridge: Cambridge University Press.

Bermúdez-Otero, Ricardo. 2015. Amphichronic explanation and the life cycle of phonological processes. In Patrick Honeybone and Joseph Salmons (eds.), *The Oxford handbook of historical phonology*, 374–399. Oxford University Press. doi:10.1093/oxfordhb/9780199232819.013.014. 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(4): 457–492.

Bloomfield, Leonard. 1933. *Language*. Chicago, IL: University of Chicago Press.

Boersma, Paul and Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic inquiry* 32(1): 45–86.

Boersma, Paul and Joe Pater. 2016. Convergence properties of a gradual learning algorithm for Harmonic Grammar. In John J McCarthy and Joe Pater (eds.), *Harmonic grammar and harmonic serialism*, 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(03): 423–441. doi:10.1017/S0954394510000177. 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(2): 333–355.

Chambers, J. K., Sarah Cummins, and Jeff Tennant. 2008. Louis Gauchat (1866-1942): Patriarch of variationist linguistics. *Historiographia Linguistica* 35: 213–274.

Cheshire, Jenny, Paul Kerswill, Sue Fox, and Eivind Torgersen. 2011. *Contact, the feature pool and the speech community: The emergence of Multicultural London English*, vol. 15. doi:10.1111/j.1467-9841.2011.00478.x.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam and Morris Halle. 1968. *The sound pattern of English*. New York, NY: Harper & Row.

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, Kyung-Ae Kim, and Hye-Kyung Kang (eds.), *Current issues in linguistic interfaces*, vol. 2, 267–294. Seoul: Hankookmunhwasa.

Coetzee, Andries W. and Shigeto Kawahara. 2013. Frequency biases in phonological variation. *Natural Language & Linguistic Theory* doi:10.1007/s11049-012-9179-z. 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 Goldsmith John A, Jason Riggle, and Alan C. L. Yu (eds.), *The handbook of phonological theory*, 401–434. Wiley-Blackwell, second edn.

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. doi:10.1177/0023830914567168. URL http://las.sagepub.com/content/early/2015/02/17/0023830914567168?papetoc.

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(1): 97–106.

Eckert, Penelope. 2008. Variation and the indexical field. *Journal of Sociolinguistics* 12(4): 453–476. doi:10.1111/j.1467-9841.2008.00374.x. 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(4): 409–438. doi:10.1016/j.wocn.2005.08.002.

Foulkes, Paul, James M Scobbie, and Dominic Watt. 2010. Sociophonetics. In William J Hardcastle, John Laver, and Fiona E Gibbon (eds.), *The handbook of phonetic sciences*, 703–754. Blackwell.

Fruehwald, Josef T. 2012. Redevelopment of a morphological class 18(1): 77–86. URL http://repository.upenn.edu/pwpl/vol18/iss1/10/.

Garrett, Andrew and Keith Johnson. 2013. Phonetic bias in sound change. In Alan C. L. Yu (ed.), *Origins of sound change: Approaches to phonologization.*, 1, 51–97. Oxford University Press. doi:10.1093/acprof:oso/9780199573745.003.0003. URL http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199573745.001.0001/acprof-9780199573745-chapter-3.

Gates, Shivonne. 2018. Why the long FACE?: Ethnic stratification and variation in the London diphthong system. *University of Pennsylvania Working Papers in Linguistics* 24(2): 39–48.

Gates, Shivonne. 2019. *Language variation and ethnicity in a multicultural secondary school.* Ph.D. thesis, Queen Mary University of London.

Gauchat, Louis. [1905] 2005. Phonetic unity in the dialect of a single village. trans. by Sarah Cummins) from l'unité phonetique dans le patois d'une commune. *Festschrift Heinreich Morf: Aus Romanischen Sprachen und Literaturen.* Halle: M. Niemeyer. 175–232. *Historiographia Linguistica* 35: 227—274.

Goldinger, Stephen D. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology* 22(5): 1166–1183. doi:10.1037/0278-7393.22.5.1166.

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* (1997): 111–120.

Guy, Gregory R. 1991. Explanation in variable phonology: An exponential model of morphological constraints. *Language Variation and Change* 3(1): 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(410): 149. doi:10.1017/S095439450000185X.

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 130years of sound change. *Cognition* 139: 83–91. doi:10.1016/j.cognition.2015.02.012. URL http://linkinghub.elsevier.com/retrieve/pii/S001002771500044X.

Hayes, Bruce. 2000. Gradient well-formedness in Optimality Theory. *Optimality Theory: Phonology, Syntax, and Acquisition* (March): 88–120.

Hayes, Bruce and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39(3): 379–440. doi:10.1162/ling.2008.39.3.379. URL http://www.mitpressjournals.org/doi/10.1162/ling.2008.39.3.379.

Herzog, Marvin I. 1965. *The Yiddish language in northern Poland: Its geography and history*. Indiana University & Mouton.

Hockett, Charles F. 1958. *A course in modern linguistics*. Macmillan.

Irvine, Judith T and Susan Gal. 2000. Language ideology and linguistic differentiation. In Paul V. Kroskrity (ed.), *Regimes of language: Ideologies, polities, and identities*, 35–83. School of American Research Press.

Jannedy, Stefanie and Jennifer Hay. 2006. Modelling sociophonetic variation. *Journal of Phonetics* 34(4): 405–408. doi:10.1016/j.wocn.2006.08.001.

Kaisse, Ellen M and Patricia A Shaw. 1985. On the theory of Lexical Phonology. *Phonology Yearbook* 2: 1–30.

Kay, Paul and Chad K McDaniel. 1979. On the logic of variable rules. *Language in Society* 8(1979): 151–187.

Keating, Patricia A. 1985. Universal phonetics and the organisation of grammars. In Victoria A Fromkin (ed.), *Phonetic linguistics: Essays in honor of peter ladefoged*, 115–132. New York: Academic Press.

Keating, Patricia A. 1988. Underspecification in phonetics. *Phonology* 275–292.

Kiparsky, Paul. 1982. Lexical morphology and phonology. In The Linguistic Society of Korea (ed.), *Linguistics in the morning calm: Selected papers from SICOL-1981*, 3–91. Seoul: Hanshin.

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 & NWAV 1994.

Kiparsky, Paul. 2006. The amphichronic program vs. Evolutionary Phonology. *Theoretical Linguistics* 32(2): 217–236. doi:10.1515/TL.2006.015. 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(3): 199–244.

Kuhn, Thomas. [1962] 2012. *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Labov, William. 1963. The social motivation of a sound change. *Word* 19(3): 273–309.

Labov, William. 1969. Contraction, deletion, and inherent variability of the English copula. *Language* 45(4): 715–762.

Labov, William. 1981. Resolving the Neogrammarian controversy. *Language* 57(2): 267–308.

Labov, William. 1989. The child as linguistic historian. *Language Variation and Change* 1: 85–97.

Labov, William. 1994. *Principles of linguistic change. Volume 1: Internal factors*. Oxford: Blackwell.

Labov, William. 2001. *Principles of linguistic change. Volume 2: Social factors*. Oxford: Blackwell.

Labov, William. 2006. *The social stratification of English in New York City*. New York: Cambridge University Press, second edn.

Labov, William. 2010. *Principles of linguistic change. Volume 3: Cognitive and cultural factors*. Oxford: Blackwell.

Labov, William. 2017. Uriel Weinreich: Builder on empirical foundations. *Journal if Jewish Languages* 5: 253–266.

Labov, William, Sharon Ash, Maya Ravindranath, Tracey Weldon, Maciej Baranowski, and Naomi Nagy. 2011. Properties of the sociolinguistic monitor. *Journal of Sociolinguistics* 15(4): 431–463. doi:10.1111/j.1467-9841.2011.00504.x. URL http://doi.wiley.com/10.1111/j.1467-9841.2011.00504.x.

Labov, William, Paul Cohen, Clarence Robins, and John Lewis. 1968. *A study of 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, Beatriz R. 1978. Where does the sociolinguistic variable stop? *Language in Society* 7(2): 171–182. URL http://journals.cambridge.org/abstract{_}S0047404500005510.

Laver, John D. M. 1968. Voice quality and indexical information. *The British journal of Disorders of Communication* 3(1): 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(10). doi:10.1111/lnc3.12147.

MacKenzie, Laurel. 2013. Variation in English auxiliary realization: A new take on contraction. *Language Variation and Change* 25(01): 17–41. doi:10.1017/S0954394512000257. 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(2): 1–10.

Mielke, Jeff, Adam Baker, and Diana Archangeli. 2016. Individual-level contact limits phonological complexity: Evidence from bunched and retroflex /ɹ/. *Language* 92(1): 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(3): 311–340. doi:10.1017/S095439451300015X.

Mohanan, K. P. 1986. *Lexical Phonology.* Dordrecht: Reidel.

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 Alessandro Duranti and Charles Goodwin (eds.), *Rethinking context: Language as an interactive phenomenon*, 335–358. Cambridge: 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(2): 171–189.

Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Joan Bybee and Paul Hopper (eds.), *Frequency and the emergence of linguistic structure*, vol. 45 of *Typological Studies in Language*, 137–157. John Benjamins.

Pierrehumbert, Janet B. 2002. Word-specific phonetics. In Carlos Gussenhoven and Natasha Warner (eds.), *Laboratory Phonology 7*, 101–139. Berlin/New York: Mouton de Gruyter.

Pierrehumbert, Janet B. 2006. The next toolkit. *Journal of Phonetics* 34(4): 516–530. doi:10.1016/j.wocn.2006.06.003. URL http://linkinghub.elsevier.com/retrieve/pii/S009544700600043X.

Pitt, Mark A., Laura Dilley, and Michael Tat. 2011. Exploring the role of exposure frequency in recognizing pronunciation variants. *Journal of Phonetics* 39(3): 304–311.

Preston, Dennis R. 2004. Three kinds of sociolinguistics: A psycholinguistic perspective. In Carmen Fought (ed.), *Sociolinguistic variation: Critical reflections*, 140—-158. New York: Oxford University Press.

Reynolds, William T. 1994. *Variation and phonological theory*. Ph.D. thesis, University of Pennsylvania.

Roberts, Julie. 1997. Acquisition of variable rules: A study of (-t, d) deletion in preschool children. *Journal of Child Language* 24(2): 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(2): 189–222. doi:10.1017/S0047404500007430.

Sankoff, Gillian and Pierrette Thibault. 1977. L'alternance entre les auxiliaires avoir et être en français parlé à Montréal. *Langue française* 34(1): 81–108. doi:10.3406/lfr.1977.4818. URL http://www.jstor.org/stable/41557797http://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(3): 275–288.

Silverstein, Michael. 2003. Indexical order and the dialectics of sociolinguistic life. *Language & Communication* 23(3-4): 193–229. doi:10.1016/S0271-5309(03)00013-2. URL http://linkinghub.elsevier.com/retrieve/pii/S0271530903000132.

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(01): 69. doi: 10.1017/S0954394509000039.

Tagliamonte, Sali and Rosalind Temple. 2005. New perspectives on an ol' variable: (t,d) in British English. *Language Variation and Change* 17(03): 281–302. doi:10.1017/S0954394505050118. 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(2): 300–336. doi:10.1075/lv.16.2.06tam. 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)* doi:10.5334/labphon.96.

Temple, Rosalind A. M. 2009. (t, d): The variable status of a variable rule. In Ōiwi Parker Jones and Elinor Payne (eds.), *Papers in phonetics and computational linguistics (Oxford University Working Papers in Linguistics, Philology, and Phonetics 12)*, 145–170. Oxford: Oxford: Faculty of Linguistics, Philology, and Phonetics, 12 edn.

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(3): 160–161.

Wagner, Michael. 2012. Locality in phonology and production planning. *McGill Working Papers in Linguistics* 22(1): 1–18.

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(6): 651–666. doi:10.1177/0261927X14528713. URL http://journals.sagepub.com/doi/10.1177/0261927X14528713.

Wedel, Andrew B. 2006. Exemplar models, evolution and language change. *Linguistic Review* 23(3): 247–274. doi:10.1515/TLR.2006.010.

Weiner, E. Judith and William Labov. 1983. Constraints on the agentless passive. *Journal of Linguistics* 19(01): 29–58. doi:10.1017/S0022226700007441. 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 Winfred P. Lehman and Yakov Malkiel (eds.), *Directions for historical linguistics*. Austin and London: University of Texas Press.