Introduction

ROBERT BAYLEY AND CEIL LUCAS

Beginning with the pioneering work of Labov, Shuy, Fasold, and Wolfram in the mid and late 1960s, the study of variation has formed one of the key areas, if not the key area, of sociolinguistics. Alone among the various sociolinguistic subfields, the study of variation has been enriched by two regular conferences – New Ways of Analyzing Variation (NWAV), now in its 36th year, and, beginning in 2001, the International Conference on Language Variation in Europe (ICLaVE) – as well as by its own journal, *Language Variation and Change*. In recent years, conferences on language variation have also been held in England (VIEW 1998, VIEW 2000), and a sociolinguistics laboratory dedicated to the study of linguistic variation has been established at Nanjing University in China. Moreover, although the study of variation began with a focus on varieties of English, French, and Spanish, variationist studies now encompass many languages ranging from Guyanese and Jamaican Creoles to Brazilian Portuguese to Chinese to American Sign Language and Australian Sign Language. Variationist approaches have also met with considerable success in studies of second language acquisition. However, despite the prominence of variationist studies in North American sociolinguistics and, increasingly, in other areas of the world, most widely used textbooks in sociolinguistics devote only a chapter or two to the study of variation. Only one textbook in sociolinguistics, Chambers (2002), is devoted primarily to variation.

This volume brings together a group of contributors widely recognized for their contributions to variationist sociolinguistics with the purpose of providing accessible overviews of the major areas of concern for students of linguistic variation. And while the chapters in this volume make it very clear that we have come a very long way in over forty years of variationist studies, they also demonstrate that three fundamental facts about variationist studies remain constant: that the variation observed in real language use is systematic and its analysis can directly inform a number of theoretical frameworks about human language use; that the development of the study of real language use has been accompanied by the development of sophisticated methods of data collection and analysis tailored to the requirements of the study of variation; and that variationist studies have very frequently received their impetus from real human situations in the areas of education, employment, and the law, and that the results of variationist studies have had very tangible and important applications in all of these areas.

In the section of the volume on theory, variation is described as it has been studied in the areas of phonology (Guy), syntax (Green, Fasold and Preston), and

style (Bell). In addition, a historical perspective on the study of variation is provided (Hazen) and variation as it pertains to historical linguistics (Montgomery) and second language acquisition (Bayley) is also examined. Finally, the role of language modality in variation is considered, with a comparison of spoken and sign language studies (Lucas).

The chapters in the methods section provide clear and comprehensive introductions to fieldwork methods for the study of variation (Schilling-Estes), to methods of quantitative analysis (Tagliamonte), and to the study of sociophonetics (Thomas).

The chapters in the applications section provide a powerful demonstration of the kind of wide impact that variationist studies can have on education (Adger and Christian, Vaughn-Cooke, Rickford and Rickford), language acquisition (Stockman), the law (Butters), and linguistic profiling (Baugh). The volume concludes with an essay by Roger Shuy, one of the founding figures of sociolinguistics, on Walt Wolfram, a scholar who has contributed directly or indirectly to nearly all of the areas covered in this book.

The important connection in variationist studies between theory, methods, and applications reflected in this volume has consistently shaped and informed the work of Walt Wolfram, to whom the volume is dedicated. All of the contributors to this volume worked very enthusiastically to produce a fitting tribute to our excellent colleague, teacher, and friend.

PART 1

Theories

1 Variation and phonological theory

GREGORY R. GUY

Introduction

The study of linguistic variation is often perceived to be quintessentially engaged with phonological phenomena. This is a manifest misperception: variationist work on morphosyntactic issues began with the original foundational articles that launched the "variable rule" framework (Labov [1969] on the English copula, and Labov [1972d] on negative concord), and continues to be among the most active areas in the field. But it is instructive to consider *why* such a misperception persists. There are two factors that drive this view. First, there exists an almost prescriptive attitude that phonology is the only domain in which linguists *should* speak of variation, arising from an uneasy suspicion that any alternations found at other levels of linguistic structure might involve intentional differences in meaning. In Labov's informal definition, variation involves "different ways of saying the same thing," and for most linguists it is easy to conclude that *runnin'* and *running* are different versions of the "same thing," but rather worrisome to make the same claim about *Kyle got arrested* and *Kyle was arrested*. Hence the view that variationists tidily confine their labors to the vineyard of phonology alleviates this existential angst about the status of morphosyntactic variation.

But a second, more interesting, reason for this view is that it is indeed quite true that work on phonological variation has been deeply intertwined with phonological theory. Phonological variation in all languages is massively structured and orderly; there is a random component, such that the surface realization of a given utterance cannot be predicted categorically, but the patterns of realizations in particular contexts are probabilistically structured with great regularity – particular realizations are strongly favored by particular phonological contexts. Most of these patterns of contextual constraints on phonological variables find clear explanation in principles of phonological organization; in other words, phonological theory can (and should) explain the variable aspects of phonology along with the categorical facts. And this relationship, as with all scientific theories, is reciprocal and reinforcing: the evidence from phonological variation has been brought to bear on a variety of theoretical questions in phonology. This includes quantitative evidence and quantitative argumentation, approaches which were historically uncommon and unfamiliar in phonological theory, but which are becoming increasingly evident in recent years (cf. for example, the work of Anttila [1997]

and Kiparsky [in press] on partial constraint rankings, and of Boersma and Hayes on Stochastic Optimality Theory [Boersma 2003, Boersma and Hayes 2001]). In this respect, work on phonological variation is comparable to the development of laboratory phonology, in that it provides new kinds of data to inform and illuminate the development of phonological theory.

This chapter explores the reciprocal, mutually illuminating relation between phonological variation and phonological theory. First, we will consider some examples of how theory contributes to explaining the data; in particular, we will see how the linguistic constraints evident in phonological variation are consistently interpretable in terms of the principles and mechanisms proposed in phonological theory. Second, we will examine some of the ways that variation data has contributed to clarifying or even resolving theoretical issues in phonology. Finally, we will discuss the general theoretical question of how to best construct a theory that models both the variable and invariant facts about the sound systems of human language, and hence explains how language can be both discrete and continuous in its organization.

Explaining the patterns: what phonological theory does for the study of variation

The fundamental observation of research on linguistic variation is that it displays, in the words of Weinreich, Labov, and Herzog (1968), "orderly heterogeneity"; in other words, the alternating variants occur in probabilistically regular patterns, not in a random distribution. These orderly patterns exhibit social regularities (e.g. higher status speakers always use more of the socially highly valued variants), which are discussed elsewhere in this volume. Our focus here is on the linguistic regularities that are also apparent. These take the form of contextual conditioning: certain linguistic contexts favor the occurrence of particular variants. Thus phonological reduction processes, if sensitive to stress, typically occur more often in unstressed syllables, assimilation processes typically occur more often word-internally than across word boundaries, and vocalization of sonorants occurs more often in coda positions than onsets. Such results are unsurprising. To a phonologist, none of the examples just cited contravenes any theoretical principle, while all of them resemble numerous cases involving categorical alternations. The central observation here is that variable processes display the same patterns of occurrence and non-occurrence that are found for categorical alternations, and hence are likely governed by the same principles and generated by the same processes of grammar. Since alternations are what phonological theories have classically been designed to account for, we can reasonably expect that extant theories incorporate explanatory principles and generalizations about linguistic structure that are relevant to variable alternations.

As an example, consider the patterns of alternation between occurrence and non-occurrence of word-final consonants. We find many cases of categorical alternation, such as French liaison, where a consonant is articulated at the ends of given words when the following word in the utterance begins with a vowel, but is absent when the next word begins with a consonant. Such cases are typically described in phonological theory as involving an underlying consonant that is suppressed under certain conditions that would be phonologically infelicitous – in this case, when it is in the coda – but retained in more favorable conditions, e.g. when it can be syllabified as an onset. In the terminology of Optimality Theory, a markedness constraint like *Coda outweighs considerations of faithfulness to the underlying form.

Parallel patterns showing the same kind of constraint but involving variable rather than categorical conditioning are also easy to find. English (also Dutch) has alternating presence and absence of final coronal stops, and this alternation is affected by whether a following word begins with a vowel or consonant, but the alternation is not categorical. That is, a word like *east* can occur as *eas'* in any following context, but the form with deletion is much more common when there is a following consonant. Thus the pattern is:

frequent, preferred:	*east end*	*eas' side*
possible, but rarer:	*eas' end*	*east side*

The generalization is that the language prefers retention before vowels and deletion before consonants. This is the same generalization that could be made about French liaison. The difference between the two cases is that in French, the dispreferred cases are absent, while in English coronal stop deletion they are not entirely absent, but simply occur less often.

This is an example of what has been described as the "stochastic generalization" relating variable and categorical observations in linguistics (Clark 2005:209, Bresnan, Dingare, and Manning 2001). Many of the principles and processes proposed in phonological theory to account for categorical facts are also evident in variable operations, in a probabilistic form. Some principle enunciated on the basis of the observation that in language A, structure X never occurs, turns out in language B to explain why structure X is very rare, although not categorically absent.

In the balance of this section, we consider some examples of how general phonological principles are reflected in the probabilistic distributions found for phonological variation. The exposition focuses on one variable which is typical of the kinds of patterns evident in variable phonology: the alternation in Brazilian Portuguese between presence and absence of word-final sibilants.

Final sibilant deletion in Brazilian Portuguese

In vernacular speech, Brazilian Portuguese shows great variation in the realization of word-final sibilants: words such as *menos* "less, minus" and *ônibus* "bus" are,

Table 1.1 *Constraints on final sibilant deletion in Brazilian Portuguese (Data from Guy 1981)*

Factor	*N*	% deleted	Factor weight
Word stress			
Stressed monosyllable	7504	6	.24
Stressed polysyllable	1375	10	.34
Unstressed	1392	53	.86
Following segment			
Vowel	3625	8	.40
Consonant	4876	16	.60
Voicing of following consonant			
−voice	2270	9	.42
+voice	2606	21	.58
Place of following consonant			
Labial	1600	14	.53
Alveolar	2240	21	.66
Velar	1036	6	.31
TOTAL	10271	13	–

for many speakers, more often realized as *meno*, *ônibu* without the final consonant. This variation is subject to a number of constraints, which are illustrated in Table 1.1.

These data raise basic linguistic questions: why do we find these patterns, and not others? Why do these contexts have the observed effects? These are the kinds of issues that phonological theory is intended to answer. Let us consider each constraint in turn.

Stress

Word stress is found to condition phonological operations and distribution in virtually every language that has a stress contrast. The direction of effect observed here is that stressed syllables have greater retention (i.e. are more faithful to underlying form), while unstressed syllables are more congenial to deletion. This is consistent with theories of prosody, positional prominence, etc., and with categorical alternations in many languages. It is also consistent with diachronic principles: in language change, stressed positions are more resistant to lenition and deletion processes.

Following segment

Increased rates of deletion in preconsonantal contexts are widely observed in variation studies. The theoretical explanation for this lies in principles of syllable structure. A word-final consonant resides underlyingly in coda position, which is universally marked and disfavored. Theories of syllable structure state this in

various ways; thus, CV phonology (Clements and Keyser 1983) treats CV as the universally unmarked syllable type, while Optimality Theory postulates NoCoda as part of the universal inventory of phonological constraints. So coda deletion is an expected repair, and a common diachronic change. However, a following vowel licenses the consonant as an onset, which is an optimal position for retention. Word-internally in Portuguese, as in many other languages, prevocalic consonants are obligatorily syllabified rightwards, as onsets. Across word boundaries, this is optional, and the outcomes are variable.

Voicing of following consonant

The data show appreciably more deletion before voiced than voiceless consonants. A theoretical explanation of this result requires one additional observation about Brazilian Portuguese. Voicing of sibilants is not phonemically distinctive in coda position; hence final sibilants assimilate obligatorily to the voicing of a following segment. The pattern shown here therefore reduces to the generalization that voiced fricatives are deleted more than voiceless ones, which has a ready explanation in markedness. Voiced fricatives are universally more marked than their voiceless counterparts; they are also typologically rarer, and raise aerodynamic problems in articulation, since the glottal impedance associated with voicing reduces the airflow required to generate the turbulence of frication.

Place of following consonant

The figures in the table indicate a robust effect of the place of a following consonant, with highest deletion rates before an alveolar, second highest before a labial, and least deletion before a velar. This is a clear example of the Obligatory Contour Principle (OCP), which states that adjacent identical elements are dispreferred. It was first proposed in phonological theory to account for the avoidance in tonal languages of sequences of adjacent identical tones, but it has been generalized to phonological processes that avoid adjacent identical segments and features (cf. Yip 1988).

As the name implies, the OCP was originally postulated to account for obligatory, categorical phenomena, but numerous gradient or variable phenomena also confirm a general preference for "contoured" sequences (where adjacent elements are dissimilar) over "level" sequences where adjacent elements are identical or similar. For example, Guy and Boberg (1997) found that English coronal stop deletion shows an OCP effect of the preceding consonant: there is more deletion after segments that are phonologically similar to the targeted /t,d/, i.e. those that share more features. Thus deletion is favored by preceding stops (e.g. *act, apt* – same in continuancy and obstruency) and alveolar fricatives (*last* – same in place and obstruency), but disfavored by preceding liquids (*cold, hard*) and labial fricatives (*left*), which share fewer features with the target.

The place data in Table 1.1 show essentially the same pattern. A conventional distinctive feature treatment of place contrasts velar, alveolar, and labial in terms of several features, as in the following matrix:

	[coronal]	[back]
labial	−	−
alveolar	+	−
velar	−	+

In this treatment, alveolar place shares one feature with labial place, but none with velar. Hence the deletion target, a coronal sibilant, is most similar in place to a following coronal consonant (like t,d,n), partially similar to labials (p,b,m), and most different from velars (k,g). The deletion facts in Table 1.1 follow this cline of similarity, implying that they are governed by a Contour Principle that is not obligatory, but probabilistic.

Constructing the theory: what variation does for phonological theory

The above examples illustrate the explanatory value of phonological theory for the analysis of variation. Now we turn to the utility of variation data for the evaluation and construction of phonological theory. As with any data, evidence of variation can be used in several ways: it can provide empirical tests of theoretical issues, it can confirm or deny the predictions of theoretical models, or it can provide facts that theory must account for. But the greatest theoretical significance of the study of phonological variation is that it has the potential to resolve theoretical issues that cannot be addressed by other means. Categorical alternations lack nuance: given a defining set of conditions, they abruptly select a single outcome. But the continuous frequency ranges of phonological variables, displaying sensitivity to a number of features of the context, offer a subtler analytical tool that can probe more finely into phonological structure. In this section I will offer an extended example of how variation data provide a unique empirical test of a theoretical issue in phonology: the treatment of lexical exceptions to phonological processes.

Phonological theory is centrally concerned with identifying generalizations about sound systems and hypothesizing mental grammatical structures that explain why and how those generalizations come about. Generative and post-generative models of phonology typically assume a bipartite architecture consisting of a phonological component, in which the generalizations are captured, and a lexicon, which lists the ungeneral, specific characteristics of individual words. For example, in the word *act*, the fact that the coda cluster /kt/ shows a constant value for the feature [−voice] throughout, and has the /k/ preceding, rather than following, the /t/, are general features of English phonology, but the fact that the vowel is /æ/ rather than /ey/ or /iy/ is one of the distinctive properties of this lexical item that distinguish it from *ached*, *eked*, and other words of English. The basic organizing principle is: general properties = phonology, specific properties = lexicon.

The problem that arises, however, is that there are many phonological generalizations that do not apply to the entire lexicon; rather, some lexical items are exceptional in certain respects when compared to most other words in the language. Thus English shows a vowel laxing alternation in *serene–serenity, obscene–obscenity*, but not in *obese–obesity*. Also, in Philadelphia English, the vowel /æ/ is tense before tautosyllabic anterior nasals and fricatives (hence tense *man*, *mansion*, *ham*, *hamster*, *half*, *after*, vs. lax *hang*, *hammer*, *planet*, *scaffold*, *have*, *that*, *sad*, *sack*, etc.); however, *mad*, *bad*, *glad* are tense despite the following /d/ (cf. lax *sad*, *Dad*, *had*, *fad*, etc.). How are such cases to be accounted for?

Although the theoretical literature on lexical exceptions has focused on categorical alternations, the same issue also arises in phonological variation. There it takes the form of lexical items that undergo certain processes at an exceptional rate, compared to other words of comparable structure. For example, the word *and* is produced without a final stop far more often than phonologically similar words like *hand* or *band*. So an adequate phonology of variation faces the same problems confronted by a categorical phonology.

Given the phonology-with-lexicon architecture, there are just two ways that lexical exceptions have been handled without dropping the generalization from the phonology. First, exception features can be attached to lexical items to co-index them with phonological processes; this is the mechanism suggested by Chomsky and Halle (1968). A lexical item that fails to undergo rule *n* can be annotated in the lexicon with a feature [−rule *n*]; similarly, a set of lexical items that undergo some rule *m* that other words do not can be annotated with a feature [+rule *m*]. Second, the exceptional outcomes can be directly represented in the underlying representation of the exceptional words, preempting the phonological processes that would otherwise apply or fail to apply.

These two approaches to lexical exceptionality have survived the theoretical shift in phonology from rules to constraint-based formalisms. Optimality-theoretic treatments of exceptionality use the same two strategies, relying either on preemptive structural marking of underlying representations or on lexically-specific constraints that apply only to co-indexed lexical items (cf. Pater and Coetzee 2005). It therefore appears that the roots of these approaches lie in the dichotomous architecture of phonology vs. lexicon – one repository for general facts, one for particular facts. The existence of exceptions implies that there are "generalizations" that are only partially true, i.e. partly general and partly specific. The dichotomy between phonology and lexicon therefore gives us two choices. We can focus on the supralexical generality of the pattern, thus retaining the phonological mechanisms that would capture it (whether they are rules, representations, or constraints), but delimiting their lexical scope by means of exception features; this is the "phonological" approach. Or, we can focus on the particularity of the exceptions by writing them directly into the lexical representations, thereby preempting the phonological mechanisms from accounting for them; this is the "lexical" approach.