1

Gregory R. Guy

VARIATION IN THE GROUP AND THE INDIVIDUAL: THE CASE OF FINAL STOP DELETION

Introduction

One of the enduring questions in linguistics has been the nature of the relationship between the individual and the group. The problem is rooted in the fact that language, while existing to serve a social function (communication), is nevertheless seated in the minds of individuals. This dichotomy is reflected many times in linguistic theory, from Saussure's *langue–parole* distinction right down to present-day argumentation in syntax, where a multiplicity of idiosyncratic dialects is invoked to account for divergence of syntactic judgments.

An important aspect of this issue is the definition of what exactly should be the subject matter of a linguistic description. Are we to write grammars of the speech of an individual, or of the language of a community of speakers? It is a necessary goal of linguistics to describe the most general object possible. However, the variation encountered at the level of "language" has led many linguists, in the pursuit of homogeneity, through successive subdivisions of the available data, until ultimately an ideal, variation-free idiolect becomes the "correct" object of attention.

It would appear, however, that the pursuit of homogeneity in this fashion is doomed to failure, because variation is to be found everywhere in language, even at the level of the idiolect. We are coming to accept variation as "inherent" in language and possibly even essential to it.

An example of such inherent variation is the process of consonant cluster simplification in English; more specifically, the variable rule by which final stops (especially /t,d/) in consonant clusters are deleted. This rule is

ISBN 0-12-432101-1

very compelling; it affects virtually all speakers of English in all but the most self-conscious styles. It is intricately conditioned, but is rarely categorical.

We conducted a study of this variable as part of a research project focusing on linguistic change and variation in the Philadelphia speech community. The project was supported by the National Science Foundation and directed by William Labov. In this chapter, I will report the results of that study. I will also describe the methodology that we developed for the study of this type of systematic variation. Finally, I will use the results to draw some conclusions about the relationship between the group and the individual. This question will be attacked through the analysis of the STRUCTURE OF VARIATION—the proportions of different variants used, the conditioning effect of different environments, the hierarchy of constraints, patterns of style shifting and social stratification, etc. Such knowledge of the structure of variation is indispensable for the understanding of historical processes of linguistic change, as well as for the synchronic study of language and its social usage.

Background

Systematic description and analysis of variation in language was inaugurated with the study of linguistic variables (Labov, 1966). It was given a formal relationship to conventional generative grammar by the introduction of the concept of variable rules—as distinct from the conventional obligatory and optional rules—in Labov (1969). Another major advance came with the introduction of probability theory and the development of computer analysis techniques by H. Cedergren and D. Sankoff in 1972. Using the Cedergren–Sankoff program one can easily estimate from data the probabilistic effects of constraints on the operation of a variable rule. The techniques used to make this estimation, however, require fairly sizable amounts of data. Therefore, most of the early studies using this program had data bases that grouped individuals together, rather than keeping them distinct (see, for example, Cedergren, 1973; G. Sankoff, 1973). This procedure was appropriately criticized by Bickerton (1973a,b) and others, since it makes it difficult to tell whether the observed variation is in fact present in the speech of individuals, rather than being simply the artifact of lumping together disparate, but internally homogeneous, individuals.

This criticism raises again our original question in a slightly different form: Is variation in the speech community the result of the diversity of the group, reflecting the organization of society into a number of discrete lects within which variation is at a minimum, or is this variation present with identical uniform structures in the speech of every individual? The truth, as it usually does, no doubt lies somewhere between these two extremes. The Baileyan "wave model" (Bailey, 1973, 1974), in which all variation is automatically viewed as a stage of a change in progress, essentially adopts the former view, considering individuals, and rule environments, to be hierarchically organized, with change gradually diffusing through contiguous environments and contiguous people. For a given speaker, variation is confined to a few environments per rule and different speakers may be ranked according to which environments are variable and which categorical. Our development of the variable rule analysis in the present study has benefited from the questions raised by Bailey and Bickerton concerning individual differences, and attempts to deal with some of these issues empirically.

A variable rule accounts for patterned variation in language by positing that every possible constraint on the rule has an associated probability as to whether or not the rule will apply. For a given instance, the overall probability of rule application is obtained by multiplying (or summing) the effects of all the constraints present in the environment, and over the long run the proportions of the different variants used in a given environment should reflect this overall probability. In this approach categorical rules (or constraints) have probabilities of 1 or 0. Linguistic change is viewed simply as a change of the values of some or all of the constraint probabilities. "Reweightings" of constraints are easily described in numerical terms. Bailey's model for the evolution of a rule, like the models considered in Labov (1972a) is also easily describable in these terms: Initially the rule has a probability of 0 in all environments; then the probabilities gradually increase, at first in a few environments, then in others, until they arrive at 1, first for the earliest environments, then for the later ones. This course could also be reversed or arrested, yielding stable variable rules which persist for a long time. However, it is clearly unjustified to assert that all presently stable variable rules are stagnated relics of earlier changing ones, especially in the case of a rule such as the one we will posit for final /t,d/ deletion. In the absence of any historical or contemporary evidence of change, such an assertion is unfounded speculation. But in any case, the advantage of variable rule methods of analysis is that they can be applied to a variety of such models of historical processes.

The Cedergren–Sankoff variable rule program estimates the probabilistic effect of each constraint on a rule, based on observations of the use of different forms of a linguistic variable. Descriptions of the use and applications of the program can be found in Cedergren and Sankoff's basic paper in *Language,* and in a number of further studies (including Griffin, Guy, and Sag, 1973; Guy, 1974). I will briefly recapitulate some of the terminology developed in those papers.

Any use of the variable rule program must be modeled on some proposed variable rule, and the data are tabulated in terms of whether that rule applied or did not apply (e.g., for a deletion rule, a rule application means an absence of the variable in a position where it could have occurred). A FAC-

TOR is any constraint on the rule that can be stated in the environmental description of the rule. (For example, if the first member of a final consonant cluster is a sibilant, this has an effect on the rate of final /t,d/ deletion, so this is one of the factors we code for.) A FACTOR GROUP is the set of all the possible factors that can occur at a given location in the environment of a variable. For example, final /t,d/ deletion is conditioned by the nature of the following segment, so we code a factor group for following segment consisting of the factors K (consonant), U (liquid),[1] G (glide), V (vowel), Q (no following segment—mnemonically "quiet"). (The capital letters are the computer codes we have assigned to the individual factors, and will be used throughout the chapter as convenient abbreviations for them.) A FACTOR VALUE is the probability associated with a factor which is found by the Cedergren–Sankoff program. An INPUT PROBABILITY for a rule is the probability that the rule will apply regardless of environmental constraints.

Final Stop Deletion: A Typical Variable Rule

Final consonant clusters in English that end in stops undergo a variable, conditioned process of simplification which we will describe as a final stop deletion rule. In running speech a speaker can leave out many such stops without producing incomprehension or evoking social opprobrium. The rule can be characterized roughly as

(1) $\$ \rightarrow \langle \emptyset \rangle$ / C ___ ##

The vast majority of the stops that fall under the scope of Rule (1) happen to be /t/s or /d/s. Relatively few words in English have final consonant clusters ending in velar or labial stops, and most of those that do are cases of /sp, sk, lp, mp, lk, ŋk/. Only the apical stops can cluster with a full range of preceding consonants, including other stops. These distributional facts make it difficult to test the full range of possible deletions implied by Rule (1), since data are comparatively sparse on the labials and velars. Therefore we will restrict our quantitative study to the examination of a rule of final /t,d/ deletion, of the following form:

(2) t,d $\rightarrow \langle \emptyset \rangle$ / C ___ ##

Although we have confined our attention to cases of word-final clusters, this rule should probably be extended to include clusters before # boundaries as

[1] We used "U" as the symbol for liquids in the following environment factor group to distinguish it from the "L" used as the symbol for laterals in the preceding environment factor group.

well, since we do find deletion in words such as *directly* and *expects*. Furthermore we have chosen to exclude from this rule any consideration of deletion of final /t,d/ after vowels. This process is very common for many black speakers, and possibly should be considered as a part of this same rule, but we found no unambiguous instances of it for any of our white speakers.

There are of course several other rules of consonant deletion in English. Standard English permits, among others, the following processes: combining two successive occurrences of a consonant into a single one (e.g., *black cat* = /blækæt/), deleting /t/ (and sometimes /p/ or /k/) from an *sts* cluster, leaving a long /s·/, as in *linguists* = /lingwis·/, and processes of assimilation as in *i's*, *tha's*, and *wha's*. Many nonstandard dialects have more extensive processes, such as VBE's final stop deletion after vowels and /l/-deletion after vocalization (e.g., *self-help* = /sef hep/). Most of these other processes also appear to be variable rules, and warrant studies in their own right.

Rule (2) as it stands incorporates none of what we know about patterned variation in the application of the rule. A number of previous studies of the rule have demonstrated that it is conditioned by such factors as whether the cluster is mono- or bimorphemic, whether a following word begins with a vowel or a consonant, *etc*. To incorporate this information, we will treat this phenomenon within the variable rule framework that has been outlined.

Studies of /t,d/ deletion have been undertaken in a variety of research contexts by Labov (1967); Labov and Cohen (1967); Labov, Cohen, Robins, and Lewis (1968); Wolfram (1969, 1971); Fasold (1972); and others. Table 1.1 summarizes the various factors that were considered in those studies and in the present one. All of the studies show considerable agreement as to the major conditioning effects on /t,d/ deletion. The conditioning factors found were the following:

1. *Grammatical Conditioning*. Many of the clusters ending in /t,d/ found in English are the result of suffixing a /t/ or /d/ representing the past tense morpheme *-ed* onto a verb ending in a consonant (e.g., *walked, fished, rained*). Such bimorphemic clusters are much more resistant to the /t,d/ deletion rule than are monomorphemic clusters (as in *expect, mist, mind*), possibly because the result of the rule would be to produce forms that were indistinguishable from present tense forms (except in the third person singular). This may be expressed as a variable conditioning on the rule according to the presence or absence of a morpheme boundary in the cluster.

A further complication is introduced, however, by those irregular verbs that have in the past tense both a vowel change AND an alveolar stop suffix, such as *tell–told, sleep–slept, leave–left,* and possibly also *go–went*. We have termed these verbs "ambiguous," referring to the ambiguity in classifying them as strong or weak verbs. They do have the inhibitory inflectional boundary before the /t,d/ (although another possible analysis suggests that they have a + boundary rather than a #), but they also have another indication

TABLE 1.1
Constraints on -t,-d Deletion: Comparison of Six Studies

	Grammatical status	Preceding environment	Following environment	Other linguistic constraints	Style	Social factors
Labov, 1967	M,P		K,V			Race
Labov & Cohen, 1967	M,P		V,~V		Casual, careful	Social class
Labov, Cohen, Robins, & Lewis, 1968	M,A,P	Obs, Son	V,~V			
Wolfram, 1969	M,P	Stop, Son, Fric	K,~K		Interview, reading	Age, race, social class, sex, racial isolation
Fasold, 1972	A,P	Fric, Son	K,V,Q	Stress		
Guy, 1974	M,A,P	S,$,N,L,F	K,U,G,V,Q	(Articulatory complexity)	Casual, careful	Geographic background

of the past tense morpheme elsewhere in the word, so that deleting the /t,d/ would not produce the communicative hazards that it would in regular ''weak'' verbs.

Other grammatical possibilities, which we have largely ignored, are derivational boundaries (such as the analysis of *past* as /pæs+t/, but see the parenthetical note in the preceding paragraph), and the possibly morphemic status of the /t,d/ in *first* and *second,* where they could be considered as reflexes of a dental suffix meaning ''ordinal number.'' In the preliminary study we did investigate the *-ed*'s that occur in past participles accompanied by auxiliaries separately from those that occur in simple past tense forms, but we found no significant difference between the effects of the two types.

The earlier studies of final /t,d/ deletion all reported that grammatical conditioning had a significant effect on the rule. Labov (1967), Labov and Cohen (1967), and Wolfram (1969) considered only the monomorphemic versus bimorphemic types. Labov, Cohen, Robins, and Lewis (1968) found a three-way distinction between monomorphemic, ambiguous, and bimorphemic clusters. Fasold (1972) treated only the past tense forms, but distinguished ambiguous and bimorphemic types. Having isolated these three major grammatical classes of clusters, we set up a factor group for our variable classes of clusters, we set up a factor group for our variable rule analysis consisting of the factors M (monomorphemic cluster), A (past tense of ambiguous verb), P (past tense of regular weak verb). The expectation was that the factor values for this group would be ordered M > A > P, that is, that M words would have a higher probability of deletion than A words, which would in turn have a higher probability than P words.

2. *Conditioning by Following Segment.* The segment (if any) that follows the /t,d/ cluster in the speech stream has been shown to have a considerable effect on the /t,d/ deletion rule. The rule is promoted by following consonants and inhibited by following vowels. (In the latter case, resyllabification frequently appends the /t/ or /d/ to the beginning of the next word, yielding, for example, /pɩk tʌp/ for *picked up.*) Thus the general effect of the rule is to produce CVC syllable structures. Consonants and vowels are known to have these same sorts of effects on other rules; for example, in most ''r-less'' dialects of English prevocalic /r/ is retained whereas preconsonantal /r/ is lost.

The finer structure of this conditioning can be examined by distinguishing—essentially on an articulatory and acoustic basis—four classes of following segments (which can be further characterized by means of distinctive feature notation), namely: consonants [+cons, −voc], liquids [+cons, +voc], glides [−cons, −voc], and vowels [−cons, +voc]. If the features [+cons] and [−voc] in a following segment favor the rule, and their binary opposites disfavor it, then glides and liquids, each having one favoring and one disfavoring feature, should show effects on final /t,d/ deletion that are intermediate between consonants (with both favoring features) and vowels (with both disfavoring features).

It is, of course, also possible to have no following segment at all, that is, silence or pause. This has no distinctive analysis in terms of the two features mentioned above, hence no obvious (or "inherent") position in the effect hierarchy of the four segmental classes. It has been treated in several different ways by previous investigators.

In the three studies in which Labov was involved, the practice was to distinguish only two categories for this environment: following vowel and following nonvowel. Thus consonants, liquids, pauses (and possibly glides?) were grouped together in opposition to vowels. Wolfram (1969) took a different approach, distinguishing consonants (presumably including liquids) from nonconsonants (which expressly included pauses). Fasold (1972) distinquished three groups—consonants, vowels, and pauses—and he reported that for his speakers pause was similar in effect to consonant. In the present study we used a factor group for following environment that consisted of five factors: K (consonant), G (glide), U (liquid), V (Vowel), Q (pause).[2]

3. *Conditioning by Preceding Segment*. The previous studies have all indicated that the preceding consonant in the cluster also influences the probability of final /t,d/ deletion. Deletion generally seems to be more probable after /s/ than after other consonants. There are, of course, a variety of ways to analyze and classify the consonants, but it appears that manner of articulation is the classificatory dimension most relevant to this rule.

Labov, Cohen, Robins, and Lewis (1968) used a two-way classification: obstruents (stops and fricatives) versus sonorants (nasals and liquids). Wolfram (1969) used a three-way distinction: fricatives, stops, and sonorants (nasals and laterals). Fasold's approach was similar to Wolfram's, except in the choice of the distinctive features to distinguish the three classes. (The other studies did not address themselves to this issue.) For the present study we adopted a five-way analysis: S (sibilants), F (nonsibilant fricatives), N (nasals), $ (stops), and L (laterals). As cases of deletion after /r/ were rare or nonexistent, we decided to consider postvocalic /r/ as being essentially a vowel for purposes of this rule. Hence no provision is made for it in this factor group and L is represented entirely by /l/ in the data. This approach has proved adequate for our present study, but not totally satisfactory, and is open to possible revision.

The above are the most powerful constraints on the /t,d/ deletion rule. Other subsidiary linguistic constraints have also been noted; for example:

4. *Stress*. Unstressed syllables are more likely to be subject to final /t,d/ deletion than are stressed ones. Fasold examined this and found it to be a

[2] It might be considered theoretically more desirable to have used three separate factor groups ([±seg], [±cons], [±voc]) to specify the following environment. This was avoided because the distinctive feature analysis is essentially derivable from the present approach, yet it requires 6 factor codes (in 3 groups) to specify only 5 classes of following environments.

fourth-level constraint. We did not consider stress in the present study, as most of our tokens are under primary or secondary stress.[3]

5. *Rate of Speech*. Probability of deletion apparently increases in proportion to the rate of speech. We did not examine this variable since we have not yet developed a simple, reliable system for measuring and coding rate of speech in natural conversation.

6. *Length of Cluster*. There appears to be a higher probability of deletion for words with triple clusters (e.g., *mixed, next, instinct, lapsed, risked, edged*), than for those with only double clusters (*mist, filled, rift, expect, mind*). However, as words in the former category are somewhat rare (and predominantly past tense verbs) we did not examine this question quantitatively.[4]

7. *Articulatory Complexity of Cluster*. The rule also appears to be affected by what could be labeled the articulatory complexity of the cluster. The measure that I developed for this is the number of changes in point of articulation that are required to execute the cluster. We would consider an /st/ cluster, with no change in point of articulation, to be easier to produce than an /ft/ or /kt/, which each involve one change. Furthermore, we would consider a /kst/ (one change) to be simpler than an /skt/ (two changes). This measure can be extended to include the point(s) of articulation of the initial consonant(s) of the following word (if any). This extended measure has a range from 0 (as in *missed out*) to 4 (as in *asked Brown*). Such a concept helps to explain why *asked* (2 changes) is overwhelmingly produced as /ast/ (0 changes) or in the metathesized form /akst/ (1 change) rather than in the full form or (if something is to be deleted) the form with a deleted /t/, /ask/ (1 change).

We analyzed some of our data using this coding system, and the results tended to confirm the hypothesis nicely. However, this is certainly a very low-level constraint, and the great increase in the complexity of the coding and computer analysis which this system required was not justified for the main body of the study by the limited increase in accuracy which it made possible.

8. *Style*. The rule is affected by style shifting, and is less likely to apply in more formal styles. Both Wolfram (1969) and Labov (1967) studied the effect of style. We have studied it quantitatively for some speakers; the results will be discussed in what follows.

9. *Social Factors*. This rule is certainly affected by social factors. The

[3] I have examined the data of several individuals for a stress effect without using the computer program, and found it to be minor in comparison to the effects described above. However, uncontrolled stress differences may account for a small portion of the differences between individuals reported in Table 1.3.

[4] This is also a minor constraint, and is partly accounted for by the articulatory complexity measure.

most thorough treatment of this aspect is Wolfram (1969). He reports effects for the following factors: age, sex, race, social class, and racial isolation.[5] Labov and Cohen (1967) also report differences between social classes. In the present study, we will discuss some age, race, and geographic background factors.

The /t,d/ Deletion Rule and the Study of Variation

Some of the implications of the /t,d/ rule have already been discussed in Labov (1972b) and elsewhere, but certain points bear repetition and expansion. As Labov notes, the rule in its most general form is "an excellent candidate for a pan-linquistic grammar" (1972b, p. 82). Written in the form

(3) $$[+\text{cons}] \rightarrow \langle \emptyset \rangle \;/\; \langle +\text{cons} \rangle \; \langle \emptyset \rangle \; \_\_ \; \#\# \begin{pmatrix} +\text{cons} \\ -\text{voc} \end{pmatrix}$$

it applies to a number of processes in languages other than English. Kiparsky has suggested that perhaps /t,d/ deletion and other such processes should not be considered as "rules of grammar" at all, but rather as "the result of general functional conditions impinging on speech performance"—an approach which would enable the linguist to avoid describing "language particular facts" about variation (1972, p. 233). However, as will be seen, there are aspects of this phenomenon that clearly defy "universal" or "functional" explanations.

Whether or not one wishes to consider final /t,d/ deletion as a rule such as (1), (2), or (3), the conditioning that it exhibits does reflect some well-accepted and compelling generalizations about language, such as that CVC syllables are more common (less "marked") than CVCC, and that full morphemes are less likely to be deleted than parts of morphemes. The phonological constraints on this process probably have a basis in the neurology and physiology of the human speech mechanism, and/or in the acoustic characteristics of speech and speech perception. (As an example of the latter, the high rate of deletion from /st/ clusters may well be partly accounted for by the spectral similarities of the [t] burst and the [s] noise.)

The grammatical constraints may indeed be based on primarily "functional" semantic considerations. Real problems of communication can arise if the past tense morpheme is frequently deleted in a system that generally places great emphasis on tense distinctions. There are many utterances where an *-ed* suffix is the only indicator of past tense and its deletion would produce a perfectly grammatical (but semantically unintended) "present tense" string. Fortunately most discourses contain multiple indications of time reference; otherwise we would often be unable to distinguish ordinary present tense forms from past tense forms with /t/ or /d/ deleted. If final /t,d/ deletion were not grammatically constrained it might severely affect the mutual comprehensibility of many speakers. Such a problem might very well arise for some speakers of vernacular Black English who have near-categorical of final deletion /t,d/ in all preconsonantal environments, regardless of grammatical status (Labov *et al.,* 1968). The semantic effect is certainly mitigated by the richness of the VBE aspectual system, but there is in such a situation great potential for structural change, for reading problems, especially with the *-ed* suffix (Labov, 1967; Summerlin, 1972), for problems of being misunderstood by nonspeakers of the dialect, *etc.*

[5] Wolfram (1971) compares VBE and Puerto Rican speakers on this (and many other) rule(s), as well as investigating in detail many of the linguistic constraints.

Structural change in the form of relexification may well have occurred for some monomorphemic words for some VBE speakers. They invariably produce such words as *test, desk* without the final stop, even producing plurals such as *tesses, desses*. For these speakers it may be correct to say that such final stops have been deleted right out of the lexicon, and that the underlying forms do not include the stop at all.[6]

Another significant aspect of this rule is the universality and uniformity of its force. It applies, to a greater or lesser extent, in virtually all native speakers of English in almost all social settings. I have noted its effect in the speech of university professors lecturing their classes, linguists delivering papers before the LSA, Nixon delivering the State of the Union message to Congress, etc. Labov (1972b) reports variable final /t,d/ deletion for 8 individuals from places as diverse as Sonora, Texas, and Lancaster, England, as well as for 11 members of a teenage Harlem peer group, the Jets. We may thus consider it quite general among all English speakers.

The "uniformity" of this rule across speakers lies not in absolute equivalence of rates of deletion in given environments, but rather in the ordering of the constraints within a factor group—for example, in the greater frequency of deletion before consonants than before vowels, in monomorphemic clusters than in bimorphemic ones, etc.

A rule of such force is a most instructive site for the study of variation in language. It demonstrates, first, that variation is inherent, and cannot be scrubbed out of our linguistic description by ever-finer subdivisions of the data. Even with 13 factors dividing the data into 75 possible environments, we find variation in 35–70% of those cells where it is possible (all filled cells except those with only one token, which must perforce exhibit either categorical deletion or categorical retention).[7] Second, as a consequence of

[6] However, even many speakers who have *tesses* as the plural of *tes'* have verb forms like *testing,* which would indicate that these final stops are still present in the underlying forms for such speakers.

[7] Furthermore, as more data are available on an individual, we find more and more cells exhibiting variation. Thus the 70% figure mentioned here is more typical of the larger data sets.

its uniformity, the figures obtained for groups (by adding together the data on individuals) are also uniform. The group norms are not just artifacts of the macrocosmic viewpoint, representing mere averages of a collection of widely scattered individual norms. Rather, they recapitulate the generally uniform norms of individuals.

I will present evidence for both of these assertions in what follows. At present, however, I should like to point out what this suggests for the study of variation.

For rules that do not exhibit significant differences between individuals, such as final /t,d/ deletion appears to be, Bickerton's criticisms of the study of group behavior (as opposed to the study of individuals) do not hold, and the "dynamic paradigm" methodology of constructing implicational matrices to model the spread of the rule through a community and through time is not particularly useful or insightful. Those methods seem to be more applicable to situations where there is independent evidence of historical or ongoing change.

Perhaps a modest taxonomy of problems in variation will be helpful. We can distinguish two relevant dimensions: (*a*) similarities and differences between individuals (within groups); and (*b*) similarities and differences between groups. (There are, of course, many different kinds of social groups which one might wish to examine—sex, age, ethnic group, *etc.*—but all of them can be accounted for by this same sort of taxonomy.)

Let us first consider cases involving the comparison of two different groups, each of which represents a sample of some speech community at a given point in space–time. In such a situation we can construct a four-celled table for comparisons of the structure of variation, shown in Table 1.2.

In Cell 1, variation is uniformly distributed throughout the community, as is partly the case for /t,d/ deletion. In Cell 2, there is variation that is different in its structure for the two different (but internally homogeneous)

TABLE 1.2
Types of Structures in Linguistic Variation

Comparing different individuals (within groups)	Comparing different groups: Similar	Comparing different groups: Different
Similar	1. Variable rule of uniform force	2. Social or geographic dialects
Different	3. Individually stratified linguistic variation	4. Combinations of 2 and 3, or true free variation

groups. Where these groups differed only in geographic location, we would have a case of dialect difference. If we were comparing different social classes, or geographically defined groups that also differed in their social makeup (as would be the case in a comparison, say, of Westchester County and New York City speakers), it could be a case of sociolinguistic variation. If we were comparing different age groups, it could be a case of sound change in progress. Certain aspects of final /t,d/ deletion would fall in Cell 2.

In Cell 3, where different groups show similar structures but there is a large variety of norms for the individuals within them, possibly even a different norm for each individual, there would be found variables that are "individually stratified," a "polylectal" situation where the group structure arises from the regular relationships among the lects. This is, of course, the situation that Bailey and Bickerton postulate for society as a whole. A typical example of this type of variation would be a post-creole continuum, where the interesting fact is not that the individuals differ, but that they can be arranged in a regular hierarchy. Of course, in a sense each lect may characterize a group in itself. The difference between such proposed groups and more conventional social groups is that the lectal grouping is impossible to define by reference to any extralinguistic social fact. Furthermore, socially defined groups are usually quite limited in number, whereas the number of possible lects often exceeds the number of individuals.

In Cell 4, one finds possible combinations of Types 2 and 3, and possible cases of true "free variation," and possible cases of individual stylistic differences mixed with intergroup differences.

Obviously, variable rule analysis of group data will only be appropriate for variation of Types 1 and 2. For Types 3 and 4, the analysis would have to begin with some treatment of individuals, possibly then scaling or grouping them in some way. The methods of implicational scaling are expressly designed for this purpose. Where data on individuals are sufficient, an alternative approach would be an initial variable rule analysis of the individuals with subsequent scaling. The constraints on the /t,d/ deletion rule show primarily Type 1 variation with some Type 2. Most residual variation between individuals is probably the product of insufficient data.

The Output of the Cedergren–Sankoff Program

The comprehensive results of our variable rule analysis of individuals using the Cedergren–Sankoff program are shown in Table 1.3.[8] For every individual, a factor value (probability that the rule will apply—or not apply—if the factor is present in the environment of a token) is given for each

[8] The appendix gives the results of a preliminary study, which is described in an unpublished manuscript by Guy.

TABLE 1.3
Factor Values for 18 Individuals

	Input	M	A	P	K	U	G	V	Q	S	$	N	F	L	Number of tokens
Philadelphians															
Vince V. 21, SP[a]	1.00	1.00	.99	.40	1.00	.73	.23	.22	.08	.88	.65	1.00	0.00	.45	201
Rose V. 47, SP	.80	1.00	.62	.50	1.00	.99	.83	.43	.14	.34	0.00	.80	.50	1.00	175
Dorothy B. 52, SP	1.00	1.00	.71	.53	1.00	.83	.51	.43	.03	.89	1.00	.73	.72	.51	260
Walt B. 21, SP	1.00	1.00	—	1.00	.89	1.00	.89	.65	.31	1.00	.35	.62	.45	0.00	145
Rose P. 58, SP	1.00	1.00	1.00	.43	.50	1.00	1.00	.85	.15	.50	1.00	1.00	—	0.00	48
Josephine P. 42, SP	1.00	.94	1.00	.53	.52	1.00	.36	.66	.43	.91	.76	1.00	.38	.94	109
Franny P. 14, SP	1.00	1.00	1.00	0.00	1.00	0.00	.55	.66	.44	1.00	0.00	.68	1.00	.61	68
Alice B. 48, CH	.99	1.00	.97	.40	1.00	.87	.53	.28	.05	.96	1.00	.65	0.00	.70	280
Joanne H. 27, CH	.93	1.00	.87	.10	1.00	.75	.57	.21	.04	1.00	.98	.72	1.00	.58	566
Karen A. 9, KP	1.00	1.00	1.00	.17	1.00	0.00	.53	1.00	.31	1.00	1.00	.44	—	.41	45
Mark W. 12, KP	.90	1.00	.34	.67	1.00	—	.93	.96	.52	.98	.71	1.00	—	.78	66
Bruce H., Jr. 10, KP	1.00	1.00	1.00	.25	1.00	1.00	.89	.52	.22	1.00	1.00	.39	0.00	.12	138
Kathy H. 8, KP	1.00	1.00	1.00	.17	1.00	—	0.00	.83	.11	1.00	1.00	.55	—	—	53
Philadelphia immigrants															
Joyce H. 32, KP	1.00	1.00	.72	.66	1.00	.34	.48	.42	.15	1.00	.33	.63	0.00	0.00	224
Bruce H., Sr. 36, KP	.79	1.00	.26	.94	.98	.55	1.00	.30	.08	.72	.48	.84	1.00	0.00	230
New Yorkers															
Chris A. 73	1.00	.60	1.00	.11	1.00	.75	.31	.34	.54	.94	.75	1.00	1.00	.78	129
Jacob S. 57	.77	1.00	.60	.41	.88	.52	1.00	.66	.85	1.00	.89	.83	0.00	.26	92
Leon A. 35	1.00	1.00	.70	.78	.46	.75	1.00	.38	.84	.44	.48	1.00	.69	0.00	95

[a] SP = South Philadelphia, CH = Cherry Hill, KP = King of Prussia.

constraining factor on which data was available. For a given environment, the probability that the rule will apply is obtained by multiplying together the factor values of all the factors present in that environment: Thus,

$$p = p_0 \times p_1 \times p_2 \times p_3 \times \cdot \cdot \cdot \times p_n$$

where p is the overall probability of rule application in a given environment, p_0 is an input probability,[9] and p_1, p_2, p_3, . . . , p_n are the probabilities associated with factors 1, 2, 3, . . . n in that environment.

The model whose formula I have just given is called the APPLICATION PROBABILITIES MODEL, as it is based on the probabilities that the rule will apply. As those who are familiar with it are aware, the Cedergren–Sankoff program also makes available for the analysis of any data set a NONAPPLICATION PROBABILITIES MODEL. In this second model,

$$p = 1 - [(1 - p_0) \times (1 - p_1) \times (1 - p_2) \times \cdot \cdot \cdot \times (1 - p_n)]$$

is the formula for the probability of rule application in a given environment. The program also supplies a chi-square measure of the goodness of fit of each model to the data. In our studies of final /t,d/ deletion we have found that the applications model almost always shows the lower chi-square value—the better fit. The fit of the nonapplications model is sometimes so poor as to obscure obvious relationships; it frequently picks out only the highest-ordered constraints. Therefore I have ignored it entirely in this chapter, and will report only the results of the applications model.

The final column of Table 1.3 reports for each individual the total number of tokens on which the results for them are based.

Analytical Methodology

The output of Cedergren–Sankoff variable rule analyses of a number of speakers will inevitably show some variety. Our results, presented in Table

[9] The input probability, p_0, is reported for each subject. As it was originally conceived, this value was intended to represent the probability that the rule would apply regardless of environment, and was supposed to serve as a collector for residual social variation. These ends have not been too well achieved in actual practice. Since p_0 is considered to be in the environment of every cell, for any set of data containing a cell showing categorical deletion, p_0 cannot be less than 1 (for the applications model). Otherwise, it would be impossible to predict accurately those categorically deleted cells, as an input probability value of less than 1 is equivalent to the statement that categorical rule application never occurs—every cell shows some retention of full forms. Since with the relatively small amounts of data we were able to obtain on our individuals it was almost inevitable that there should be several 0 out of 1 cells in every data set, the input probabilities are almost always at 1.00. Where they are not, it is more a measure of the largeness of the data set than the "overall pattern" of the speaker. The only way to overcome this problem is to have very sizable amounts of data in every cell.

1.3, are for certain factors very diverse. This presents several problems. On the practical side, there is the problem of trying to reduce this mass of numbers into a manageable, comprehensible form. More substantively, we want to know what it all means, how much of this diversity reflects actual differences between speakers, and how much is due to mere statistical fluctuation and smallness of sample size. Finally, we want to know how well the individuals mirror the behavior of the group, and vice versa. Answering this range of questions will require a variety of techniques. We have developed several which have proved to be very useful.

The first thing we had to know about the /t,d/ deletion rule was what regularities existed in the relationships between factors within factor groups (i.e., whether the value of M was always greater than the value of P, K always greater than V, etc.). This information was obtained from a pairwise comparison chart for all possible intragroup factor comparisons, which is prepared as follows (see Figure 1.1): A "1" is entered in the appropriate cell wherever the factor at the top of the column is GREATER than the factor to the left of the row. If two factors are equal, a "–" is entered in the relevant cell below the diagonal. Thus a speaker who had values K > G > U > V = Q would be entered as illustrated in Figure 1.1. In Figure 1.1, the K (at the top of the chart) is greater than all other factors, hence all the cells below it have a "1" entered. The U is greater than only V and Q, so only those two cells in the second column have "1" entered. Since V and Q are equal, we enter a "–" at the bottom cell of the fourth column, and so on. After all these comparisons have been entered for all speakers, we order the factors according to the most regularly observed pattern, so as to produce a chart with the maximum number of marks below the diagonal. The result for this study is shown in Figure 1.2. For the ordering shown in Figure 1.2, all the cells below the diagonal represent the preferred relationships, and any marks in the other half of the chart are deviations from the majority pattern. Furthermore, the farther a deviation is from the diagonal, the more deviant it will be considered.

	K	U	G	V	Q
K					
U	1		1		
G	1				
V	1	1	1		
Q	1	1	1	–	

FIGURE 1.1 *Sample pairwise comparison chart.*

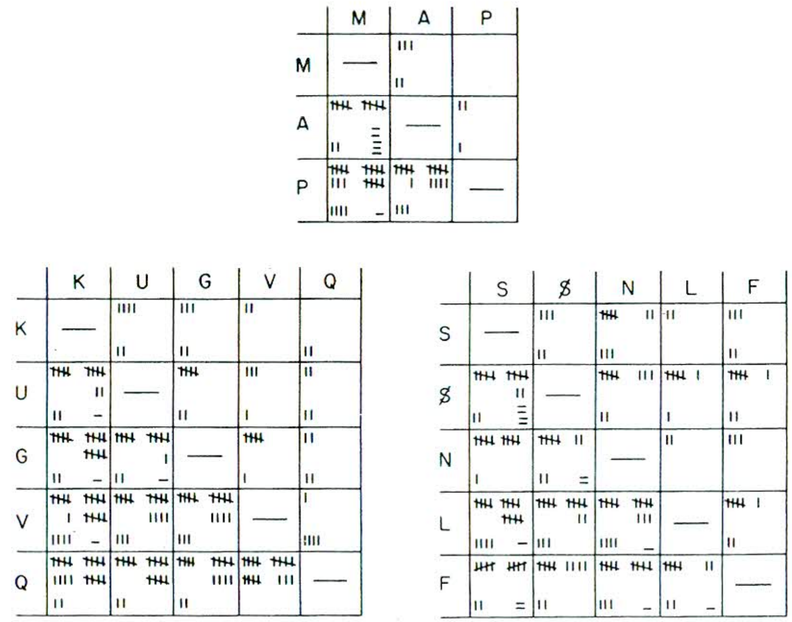

FIGURE 1.2 *Pairwise comparisons of factors. New Yorkers coded in lower left-hand corner of each cell.*

From such a display a summary of the pairwise comparisons can be obtained, as is shown in Table 1.4. The three numbers in each cell represent the number of people who had the factor written above the cell respectively greater than, equal to, or less than, the factor to the right of the cell. On the basis of these two displays we can establish a very strongly ordered relationship of M > A > P, and of K > U > G > V > Q. For the preceding environment factor group, a more weakly ordered relationship is found, with much larger numbers of deviations: S > $ > N > F > L (although the positions of N and F are somewhat questionable, and must be further examined).

There are several other obvious facts to be noted about Figure 1.2 and Table 1.4. In Figure 1.2 we observe the striking fact that most of the "deviant" relationships in the Q column are found for New Yorkers. This implies a dialect difference, which will be discussed further. Also in Figure 1.2 we note the large number of individuals who have M equal to A. We will discuss this "high-A dialect" at some length. Finally, in Table 1.4 it should be observed that although the ordering of a factor with respect to its immediately adjacent factors is not always very strong, ordering with respect to the second or third factor above or below it in the hierarchy is always quite good.

The question that now arises is, how many of the observed deviations from the regular (i.e., majority) pattern are due to small amounts of data on a

TABLE 1.4
Pairwise Comparison of Constraints on -t,-d *Deletion for 19 Philadelphians and 4 New Yorkers*[a]

	M	A
A	10–5–3 2–0–2	
P	18–1–0 4–0–0	16–0–2 3–0–1

	K	U	G	V
U	12–1–4 2–0–2			
G	15–1–3 2–0–2	11–1–5 2–0–2		
V	16–1–2 4–0–0	14–0–3 3–0–1	14–0–5 3–0–1	
Q	19–0–0 2–0–2	15–0–2 2–0–2	17–0–1 2–0–2	18–0–1 0–0–4

	S	$	N	F
$	12–4–3 2–0–2			
N	10–0–7 1–0–3	7–2–8 2–0–2		
F	10–2–3 2–0–2	9–0–6 2–0–2	10–0–3 3–1–0	
L	15–1–2 4–0–0	12–0–6 3–0–1	13–1–2 4–0–0	6–1–7 2–0–2

[a] Top line in each cell represents Philadelphians; bottom line in each cell represents New Yorkers. Each group of three numbers shows respectively the number of individuals who had the first factor greater than, equal to, and less than the second factor for that cell.

particular factor? For example, suppose we had a deviant value for F that was based on a data set that included plenty of data on all the other factors, but only 5 tokens of /ft/ or /vd/ clusters. The deviance of such a value would very likely be a result of insufficient data, and one would attribute much less significance to it than to a deviant F value that was based on, say, 35 tokens.

We can begin to answer this question with a version of the pairwise comparison chart that also includes a third dimension for number of tokens. This is done in Figure 1.3, which shows the comparisons of factor pairs plotted against the smaller number of tokens for the pair. For any pair x–y, a "1" is entered if $x > y$, a "–" is entered if $x = y$, and a "0" is entered if $x < y$. For example, if an individual had K > V, with 52 tokens of K and 35 tokens of V, a "1" would be entered in the K–V row at the 35th column (since, presumably, no comparison is stronger than its weaker link).

From Figure 1.3 it is immediately obvious that most of the deviations

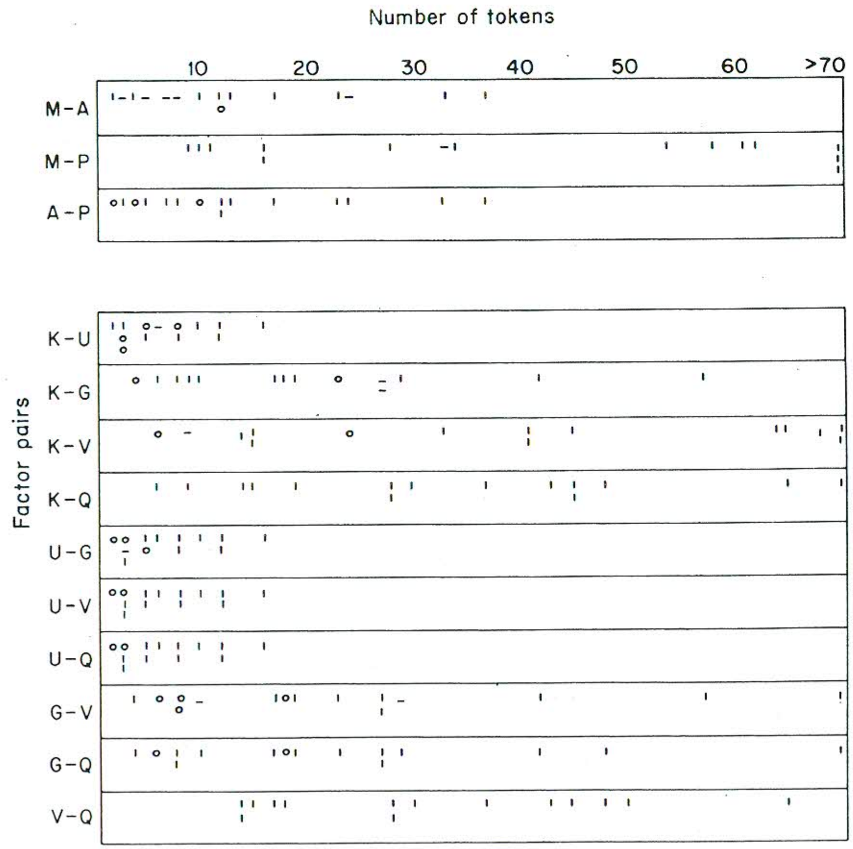

FIGURE 1.3 *Pairwise comparisons of factors by data quantity.*

TABLE 1.5
Percentage Distribution of Expected Orderings in Figure 1.3

Factor group	Number of tokens 0–5	6–15	16–35	36+
MAP	50%	73%	85%	100%
KUGVQ	56%	81%	80%	100%
N	8	15	13	9
	27	48	36	31

from the majority pattern for the grammatical status and following environment factor groups occur in the range below 10 tokens. Above 10 tokens there is 90% comformity with the expected pattern, whereas below 10 tokens only 63% of the relationships are as expected. Above 35 tokens, there is 100% conformity.[10] The percentage distribution of deviations is shown in Table 1.5.

These observations clearly demonstrate the inadequacy of statements based on small amounts of data. For this particular rule, an acceptable, reproducible level of accuracy is not obtained until each factor has at least 10 tokens representing it. Furthermore, each factor should be represented by, say, 4 or more cells; it would not be too useful to have 20 tokens on a factor, but all crammed into a single cell.

There are three ways to meet these requirements. First, one can limit the number of factors analyzed, concentrating on what appear to be the major effects, and ignoring (or preferably, combining with other factors) those factors with minimal amounts of data. Second, one could simply try to obtain sufficiently large volumes of speech from each informant. Most interviews reported for sociolinguistic surveys to date have been between 30 min and 1 hour long. For a good study of /t,d/ deletion, interviews of at least 2 hours are required. In our field work in Philadelphia, Payne developed techniques which yielded interviews of up to 3 or 4 hours in length, providing sufficient data for individual studies of a wide variety of linguistic problems (Payne, 1974).

A third way of obtaining sufficient data for a variable rule analysis is of course to lump together the data for several people. In some studies this will be the only possible way of obtaining sufficient data for a valid analysis. Hopefully the results reported here will shed some light on the question of

[10] For the preceding environment factor group, some deviations occurred at substantially higher levels of data quantity. This is probably due to the tertiary nature of these constraints and the smaller average differences between factor values in this group.

when such a procedure is appropriate, and when one of the other two measures will be required.

Another useful and instructive way to look at the data is the individual accuracy chart, shown in Figure 1.4. The chart provides a more global measure of the behavior of individuals and its relation to overall data quantity. In Figure 1.4 each individual has been plotted according to the total number of tokens in his or her data set and the number of "correctly" ordered factor

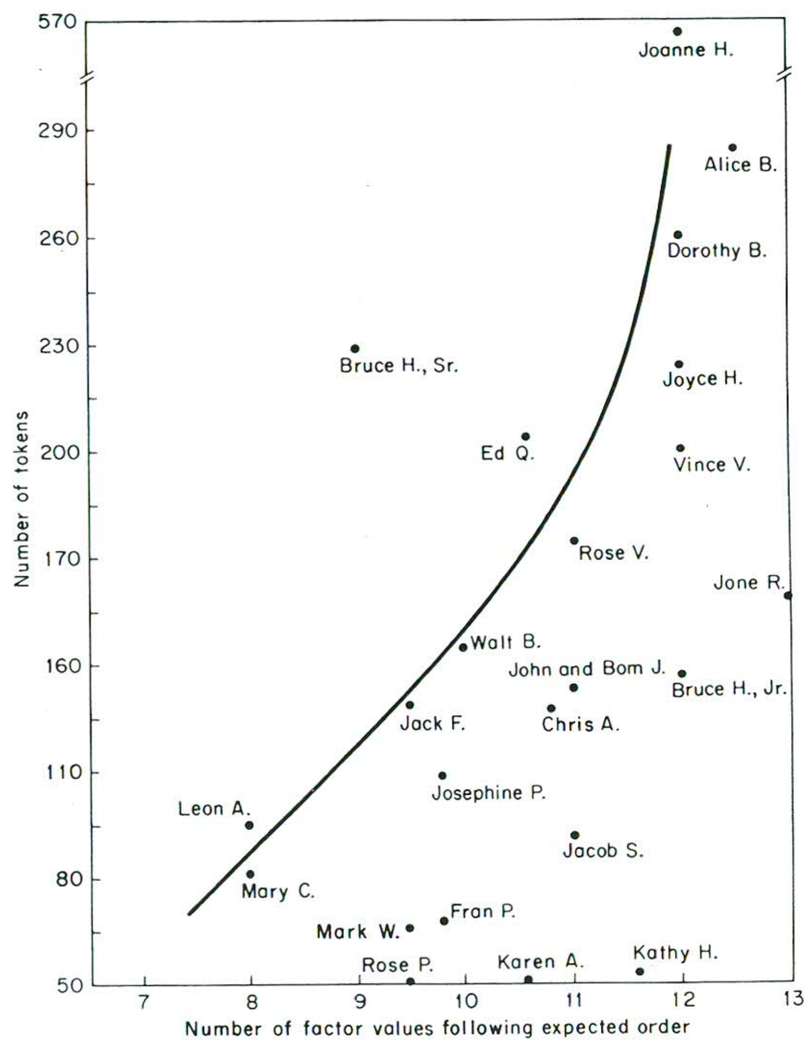

FIGURE 1.4 *Individual accuracy chart.*

values found in his or her variable rule program results.[11] (By "correctly ordered" I mean essentially "corresponding to the majority pattern.") For example, Joyce H. has 3 correctly ordered values in the MAP group, 4 in the KUGVQ group (U is too low), and 5 in SN$FL (the equivalence between F and L is not considered to be an incorrect ordering) for a total of 12 (see Table 1.3 for the numbers). Hence she is plotted on Figure 1.4 at the point for 12 correct values and 224 tokens.

Figure 1.4 clearly indicates a trend for improved reliability in factor values with increasing amounts of data. On the basis of such a trend we should be able to draw a line indicating the minimum level of accuracy to be expected in a data set of any given size. The curve drawn in Figure 1.4 may be taken as a rough approximation of such a line. Anyone above or to the left of that line would represent a significant deviation from the general pattern; there would be more deviations in their results than could be accounted for by mere random fluctuation, given the quantity of data available. It is interesting and instructive to note that the only people to the left of the line drawn in Figure 1.4 are non-Philadelphians, who presumably have different dialectal norms which account for some of their "deviations." (The worst level of "accuracy" [i.e., the maximum deviance] for the largest amount of data is Bruce H. Sr., who has the most eclectic dialect history of any of our subjects, having spent various portions of his youth in northeastern Pennsylvania, Newark, Philadelphia, and King of Prussia.)[12]

We have seen, then, that the individual accuracy diagram is a useful display which provides a quick way of estimating whether or not deviations from the norm in a set of factor values may be due to random fluctuation and insufficient data. But it must be borne in mind that this is a rather coarse measure. It is not corrected for any of the facts that might affect the significance or insignificance of the deviations, such as the quantitative magnitude of the incorrect orderings, which points are incorrectly ordered, and the amount of data on which an incorrectly ordered point is based. Frequently we find a set of "deviations" that are in a common direction and may have a linguistic explanation. Some of these have been corrected for in Figure 1.4,

[11] The following procedures were used to arrive at the values shown in Figure 1.4: Factors whose values were equal (to within ±.03) to that of a factor adjacent in the expected order were eliminated from consideration (they were considered neither "correct" nor "incorrect"); where an individual had less than 13 factors on which to base the abscissa of Figure 1.4, their number of correct points was adjusted upwards to reflect a base of 13 factors; to determine the number for the preceding environment factor group, only a partial ordering of the factors was used—$ was considered only in relation to S; L and F were not compared to each other; this was done as a partial correction for the tertiary nature of these constraints.

[12] Dialect history alone, however, probably does not account for the number of reversals of expected factor orderings found in Bruce H.'s results. Two of his deviant factors, A and F, had very small numbers of tokens. His results as a whole are somewhat unreliable in that the program did not "converge" (i.e., did not actually reach a final solution for a "best possible fit").

but others, such as the high F value which accounts for Joanne H.'s one incorrectly ordered point, have not been. The fact that the relationships in the MAP and KUGVQ factor groups are much more stable than those in the SN$FL group is not taken into account. This display of the results can be very useful, but one must be aware of its limitations.

So far our analysis has been strictly qualitative. Now we shall take a quantitative look at the actual numbers involved. Two types of graphic display suggest themselves. First, we can continue the pairwise comparisons developed in Figures 1.2 and 1.3 but using the numerical difference of the two values instead of just noting which is the larger. Plotting the difference against the smaller number of tokens for the pair (just as in Figure 1.3), we obtain a display such as those shown in Figures 1.5, 1.6, and 1.7. These quantitative factor pair charts should show values scattered over a wide range for small amounts of data (because of random fluctuation), but a narrowing down of the range for larger amounts of data, if what we are examining is at all stable. Generally speaking, the scatter of points should fall under a bell-shaped curve, whose peak would presumably represent the true behavioral norm, the ideal difference between the two factors being compared.

Figures 1.6 and 1.7 show such a pattern very nicely. For Figure 1.6, which compares the G and V factors, the width of the range is 1.48 below 20 tokens but only .73 above 20 tokens (and 7 of the 8 people above 20 tokens lie within a range of only .39). Similarly for Figure 1.7, which compares K and Q, the range width below 30 tokens is .8, whereas above 30 tokens it is only .16. What we also expect for such charts is that the individual values at the peak for maximum numbers of tokens should correspond closely to the values obtained from the amalgamated data for the whole group. This is the

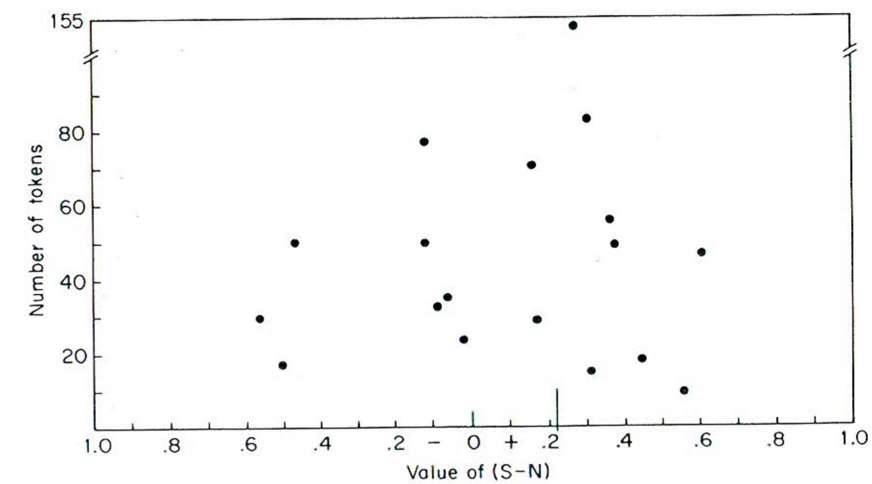

FIGURE 1.5 *Distribution of S–N values by data quantity.*

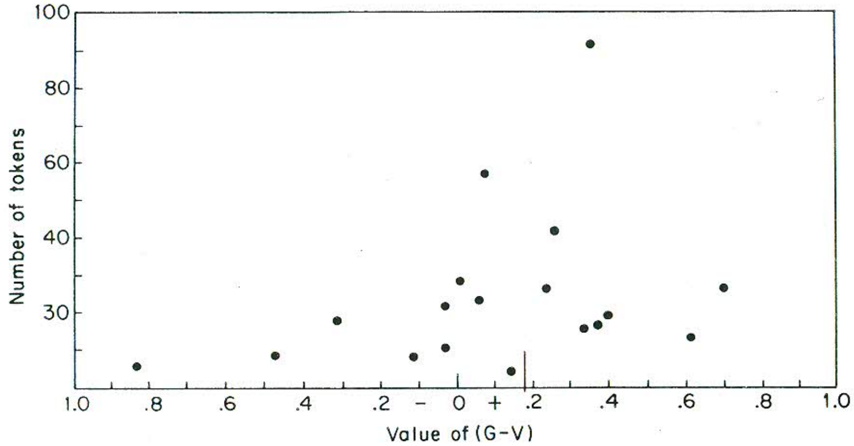

FIGURE 1.6 *Distribution of G–V values by data quantity*

case for Figures 1.6 and 1.7, where the vertical line indicates the group value. These constraint orders, K > Q, G > V certainly appear to be among the more stable relationships found for this rule.

The case is not as clear for Figure 1.5, however. There the comparison is for the factors S and N, which are not as "strictly ordered" as the pairs considered in Figures 1.6 and 1.7. The entire preceding segment factor group shows much greater variability and many more "incorrect orderings" than the other two factor groups, as can be readily seen in Figure 1.2 and Table 1.4.

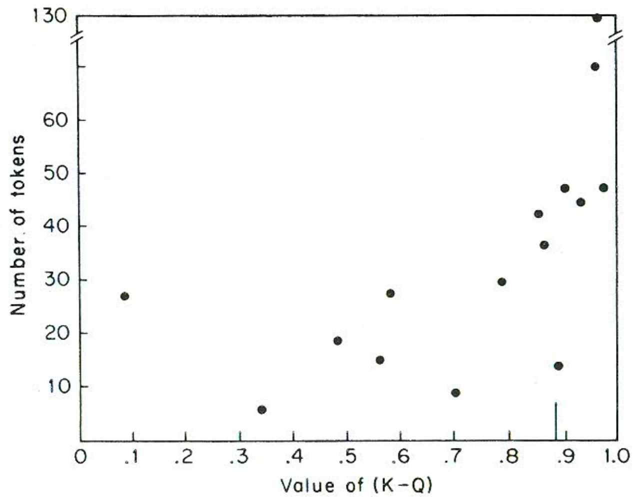

FIGURE 1.7 *Distribution of K–Q values by data quantity.*

Of course, this might be expected in a language such as English, where there is a very general tendency toward processes of anticipatory assimilation, and preceding environment is usually less significant for phonological rules than following environment. Because this factor group has less effect on the rule than the others, there tends to be a smaller average difference between factors in this group, which would produce more "overlapping" (apparent reversals of order) due to random fluctuation alone. All this is reflected in Figure 1.5, where the narrowing and peaking of the distribution occurs only above 50 tokens, if at all. It is possible that the group data in Figure 1.5 represents only an average value, rather than a behavioral norm toward which everyone should tend if we had sufficient data on them. It might be the case, however, that even Figure 1.5 would begin to peak if we had enough people with large amounts of data.

Another, perhaps more obvious, way to display the quantitative data is simply to plot all the individuals' values for a given factor against the number of tokens. This is done in Figure 1.9 for the P factor and Figure 1.8 for the A factor. If people are randomly fluctuating around a central value, the scatter of points in such a chart should also fall under an ordinary bell-shaped curve. This is essentially the case in Figure 1.9, with the notable exception of Joanne H., who clearly has a different norm for this factor. The other values shown on the chart, however, show with increasing numbers of tokens a convergence of the P value on a figure between .4 and .5, which is precisely where the P values for the group data fall. The scatter of points in Figure 1.8, however, is obviously quite different. There is clearly a bimodal distribution of values for the A factor, indicating two discrete norms for treating these

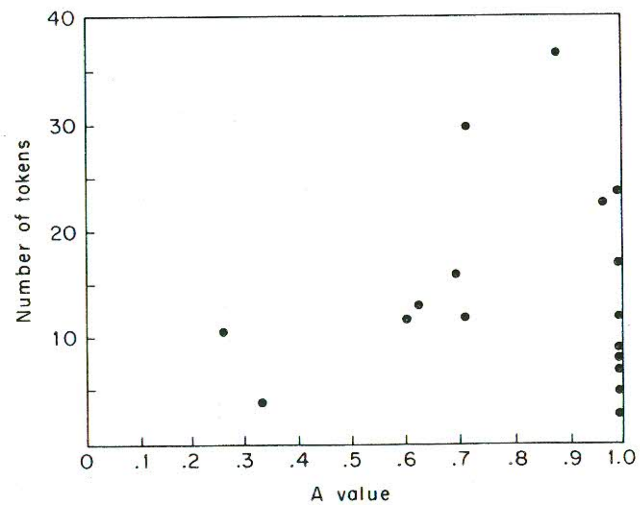

FIGURE 1.8 *Distribution of A values by data quantity.*

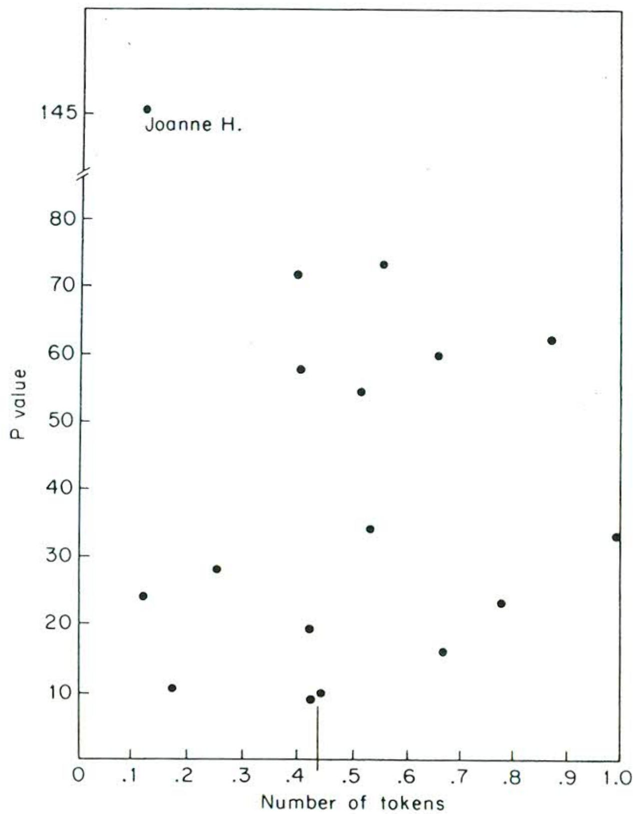

FIGURE 1.9 *Distribution of P values by data quantity.*

"ambiguous" verbs. It is significant that we find this bimodal distribution for precisely that factor for which our linguistic analysis suggests two possible interpretations. This pattern will be further discussed.

In using all these graphical methods, our goals are both methodological and substantive. As a practical matter, we wish to discover how much data is required in order to achieve a reasonable degree of reliability and accuracy in the factor values. In this regard, the figure of 30 tokens per factor seems to be an appropriate goal if reliable results are to be obtained.[13] But additionally, we wish to establish certain things about the relationship between the individual and the group: what the predominating relationships among factors are, how generally these are found in the speech of individuals, what central values there are, if any, toward which factor values for both groups and individuals tend when enough data is available, and whether there are differences between groups, such as social class or dialectal patterns.

[13] It is interesting to note that $N = 30$ is the approximate dividing line between "large" and "small" samples in statistics; below this figure different tests of significance (such as the t-test) must be used.

Analysis and Discussion

Social and Geographic Dialects

One of the most striking facts about Table 1.4 is the difference in the position of the Q factor between the New Yorkers and the Philadelphians. All 4 of the New Yorkers have Q greater than V; this is true of only 1 of the 19 Philadelphians. Two New Yorkers have Q greater even than K, whereas not one Philadelphian does. This is a clear example of a genuine dialect difference (Type 2 in the taxonomy presented in Table 1.2). For the New Yorkers, a following pause is like a following consonant in its effect on final /t,d/ deletion: It promotes it. This finding essentially corroborates Labov's practice in the original two New York studies of grouping Q together with K in opposition to V.[14] For the Philadelphians, however, Q is an extremely conservative environment, even more so than a following vowel.[15] (This can also be seen from Figure 1.7 which shows the amount by which K exceeds Q for the Philadelphians—for most of them this amount falls between .7 and .96, out of a maximum possible 1.0.) The group figures summarize this dialect difference:[16]

	K	V	Q
3 white New Yorkers	1.0	.56	.83
9 white Philadelphians	1.0	.38	.12

Let us consider the significance of this finding for linguistic theory. In all our studies, K and V are rigidly ordered, as are U and G where data on them is sufficient. The distribution of Q values, however, is highly variable, and to a certain extent bimodal. This tends to strengthen Labov's suggestion that the consonant–liquid–glide–vowel constraint hierarchy might be a universal one. The pause, however, is physically, acoustically, and functionally outside the K–U–G–V system, and therefore is susceptible to differing analyses by different speakers or dialects. Kiparsky's suggestion that observed patterns of variation might be accounted for by universal functional conditions

[14] However, Labov's New York City speakers were black, and there is evidence for a VBE pattern for this rule distinct from the SE pattern of middle-class whites, as will be discussed.

[15] This low-Q pattern was also found in a study of 8 white speakers from the southern and southwestern United States.

[16] The figures for the group of 19 Philadelphians illustrate the same point:

	K	V	Q
19 Philadelphians	1.00	.40	.19

But this group includes two black speakers (Johnny and Bom J.) who have Q values that are very high, although still the lowest in the factor group:

K	V	Q
1.00	.79	.75

is clearly untenable for a case such as the Q factor. Though it may one day be possible to explain the K–U–G–V pattern by reference to the sort of conditions Kiparsky has in mind, the effect of pause is arbitrarily defined for a given dialect—it must be learned by children acquiring a dialect, and must be accounted for in the grammar of a dialect. Clearly it is impossible to write "descriptively adequate" grammars that pay no attention to variation.

These facts about the dialectal differences in the effect of pause may account for the different treatments of it by Labov and Wolfram. As I have noted, Labov treats Q as being equivalent to K, whereas Wolfram treats it as being equivalent to V.[17] Fasold treats the three separately, and reports that K and Q are roughly equivalent. It may well be that Wolfram's speakers were predominantly of the low-Q type like the present sample of Philadelphians. However, the high-Q pattern appears to be typical of black speakers as well as of New Yorkers. Fasold's black Washingtonians, the black New Yorkers studied in Labov, Cohen, Robins, and Lewis, and the black Philadelphians that we have studied (cf. Note 16) all typically show high rates of /t,d/ deletion before a pause. If this is a general feature of Black English, Wolfram's speakers would be unusual in their divergence.

Uniformity

Table 1.6 shows the group values for

1. Seven South Philadelphians: Vince and Rose V., Dorothy and Walt B., and Josephine, Rose and Franny P.
2. Nine Philadelphia adults: the 7 people in Group 1 plus Joanne H. and Alice B.
3. An earlier "cosmopolitan" sample of 15 individuals from a variety of dialect regions, including some Philadelphians (Jack F., Jane R., Johnny and Bom J., and Mary C., none of whom appear in Groups 1 or 2)
4. Nineteen Philadelphians, including all the people in Group 2, all the Philadelphians in Group 3 plus 4 children and 1 adult from King of Prussia (a Philadelphia suburb—see Payne, 1974): Bruce H. Jr., Kathy H., Mark W., Karen A., and Joyce H.

Also in Table 1.6 we have repeated the figures for the one individual for whom we have very large amounts of data, Joanne H.[18] (Her data are also included in Groups 2 and 4.)

[17] These scholars presumably based their decisions on preliminary inspections of the data, which could possibly have been in error.

[18] The large quantities of data on Joanne H. and Alice B. were obtained by Payne during a study which was designed to collect on tape a substantial body of natural speech. Payne's method was to accompany the subject throughout most of a normal day, taping virtually everything the subject said in that time. This material was tremendously valuable for the present study, and has many other possible uses as well.

TABLE 1.6
Comparison of Group and Individual Factor Probabilities

	Input	M	A	P	K	U	G	V	Q	S	$	N	F	L	Number of tokens
1. 7 South Philadelphians	.80	1.00	.78	.51	1.00	.93	.58	.55	.20	1.00	.68	.88	.62	.67	1009
2. 9 Philadelphia adults	.86	1.00	.82	.39	1.00	.86	.56	.38	.12	1.00	.84	.78	.69	.66	1860
3. Cosmopolitan sample—15 speakers	.79	1.00	.91	.49	1.00	.89	.60	.36	.53	1.00	.86	.65	.44	.57	1931
4. 19 Philadelphians	.87	1.00	.97	.44	1.00	.77	.59	.40	.19	1.00	.81	.76	.46	.62	2886
5. Joanne H.	.93	1.00	.87	.10	1.00	.75	.57	.21	.04	1.00	.98	.72	1.00	.58	566

The close correspondence among the sets of figures in Table 1.6 is further indication of the striking uniformity of this rule. There is total unanimity as to the three factors that maximally promote the rule—M[onomorphemic], following K, and preceding S. For the two factor groups that are generally considered to contain the "highest order" constraints on this rule—grammatical status and following environment—there is no overlap whatsoever in the ranges of factor values within a group (except for the high-Q value in Group 3, which included high-Q speakers). The values for K–U–G–V in all data sets indicate that the [±cons] feature is unanimously given greater weight than the [±voc] feature. Only in the third factor group—preceding environment—is there much overlap in the ranges of factor values. But despite this overlapping, there is good agreement as to the ordering of the factors in the S$NFL pattern, excepting only that factor on which we have the least amount of data, F. Although the evidence from the individual analyses suggested that F should be placed between N and L in this hierarchy, only one of the group analyses shown in Table 1.6 shows that order. The other four factors in this group all follow the expected S–$–N–L order for all five data sets in Table 1.6, except for the reversal of $ and N in the first data set.

Besides F and Q, the only factors that show relatively wide ranges of values are P and V. These are respectively the most and second most conservative environments in their respective factor groups. It seems that the range of values that they display may be reflecting social class differences between the speakers. Apparently the most socially marked environments for deletion are the most conservative ones; middle class "respectable" speakers correct their /t,d/ deletion mainly by tending toward categorical retention in past tense and prevowel and prepause environments, while leaving their behavior in other environments unchanged. If such is the case the mark of a "proper" speaker will be a large difference between the factor values for the most and least conservative environments.

The high degree of correspondence between the group values and the individual values of Joanne H. is further evidence that speakers have essentially identical norms for final /t,d/ deletion, and that one only needs to obtain large amounts of data to demonstrate this fact. We feel that for most of the constraints on this rule (and other cases of "Type 1" variation) differences between individuals are primarily random perturbations due to paucity of data rather than real behavioral differences.

These results shed some light on some of the theoretical issues about variation that were discussed in the beginning of the chapter. As I mentioned, Bickerton and others have objected to the practice of using group data in studying variation. These writers assert that much of the "apparent variation" found in several variable rule analyses by other researchers using group data is the result of an inappropriate lumping together of different speakers and different conditioning environments. The primary orientation of this "dynamic paradigm" is toward producing "polylectal grammars" whose rules are categorical, or variable in at most only one or a few environments per speaker per rule. For a linguist working in this paradigm, persistence of massive variation is an indication that another conditioning factor or another way of dividing up the data must be found. As I have already suggested, such a search for the lect free of variation is inappropriate for many important areas of linguistic research. Further, the results of my investigations of data quantity indicate clearly that to subdivide the data too finely—by limiting the scope to individuals and multiplying the number of environments—is inherently self-defeating. Patterns and regularities in the data are obscured by such a procedure, rather than revealed. Such a procedure maximizes error as it minimizes cell size. This effect is further magnified by the practice (or requirement) of the dynamic paradigm that each different environment must be considered unitary (as they are each accounted for by a separate rule), rather than as the product of several independent effects (as is the case when a single variable rule is used to describe all environments). In the former case, a given item of data contributes information only about the cell it resides in, whereas in the latter case a datum can provide information about all the conditioning factors in its environment.

The High-A Lects

There is one factor in our data for which a study based on groups rather than individuals would be misleading, in just the way that Bickerton has suggested. The data on the A factor shown in Figures 1.2 and 1.3 are unusual in that a large number of people have the value of this factor equal to the M value (with both A and M equaling 1.00). A few even have A higher than M. Figure 1.8 shows the distribution of A values by numbers of tokens. It clearly indicates a bimodal distribution—one group clusters around .70; the other puts A at 1.00. Increased numbers of tokens have no effect on this pattern.[19] (Such a bimodal pattern for this factor was first suggested by the figures for four speakers reported in Labov, 1973.)

For this factor, the group values shown in Table 1.6 clearly do represent mere averages of distinct norms. This is the sort of situation where analyses of individuals are quite necessary. But there are indications that the high-A lect may be a social class and language acquisitional pattern, just as the high-Q pattern is a feature of geographic (and possibly ethnic) dialects. If such is the case, it might be possible to redefine homogeneous groups on the basis of extralinguistic facts.

[19] The fact that Joanne H., at .87 and 37 tokens, is rather lower than the other high-A speakers does not obscure this bimodal distribution. There is still a gap of .15 between her A value and the highest value in the low-A group (the .72 of Joyce H.) into which no one falls.

The linguistic significance of the two norms for the A[mbiguous] category is clear. The group with the lower value for A has apparently perceived the ambiguity of these /t,d/-final words, and retains them somewhat more than monomorphemic /t,d/'s (since they represent the *-ed* suffix), but somewhat less than those in regular past tense verbs (since the past tense morpheme is encoded elsewhere in the word). Alternatively, we could say that these speakers have analyzed the A words as having a + boundary before the /t,d/ and have assigned this boundary a restraining effect on the rule which is about half that of the # boundary in regular verbs.

The group with a high-A value, however, has apparently begun to obliterate this irregular class of verbs, and thrown its contents in with the ordinary "strong" (i.e., ablauting) verbs. Thus *sleep, tell, leave* are identified as vowel changing verbs with past tense forms *slep', tol', lef'* just like *swim–swam*. For some of these verbs there are even direct analogies with other strong verbs; for example *feed–fed* has the same vowel change as *sleep, keep, leave*. When the irregular verbs are analyzed in this way, their final *t*'s and *d*'s are either deleted preferentially or are not even entered in the dictionary form. In the latter case they would not actually undergo deletion because they are not there to begin with.[20]

A preliminary examination of the social distribution of the "high-A" dialect reveals two promising avenues for further research. First, the tendency to delete /t,d/ preferentially in the A category seems to be stronger in the children in our sample, as can be seen from the group figures for the four King of Prussia children studied:

	M	A	P
4 King of Prussia children	.96	1.00	.35

Second, among adults this pattern appears to be a characteristic of working class speakers, and is possibly socially stigmatized. Consider, for example, the four individuals in our sample who are probably lowest in socioeconomic status, Chris Andersen of New York City and the three members of the Porta family of South Philadelphia:

	M	A	P
Chris A., 73	.60	1.00	.11
Josephine P., 42	.94	1.00	.53
Rose P., 58	1.00	1.00	.43
Fran P., 14	1.00	1.00	.00

[20] For testing this hypothesis more fully it is sometimes useful to examine the nonapplications model provided by the Cedergren–Sankoff program. There we find that for most of those individuals for whom the applications model gives M = A = 1.0, the nonapplications model reports an A value even HIGHER than the M value. This indicates that for these speakers the ambiguous words are either subject to the deletion rule even more frequently than monomorphemic words, or else have relexified underlying forms that do not contain a final /t,d/.

If these tendencies are confirmed by further investigation, they would suggest the following: Among middle class children, the high-A lect may be a stage in language acquisition; after a child has learned that there are vowel-changing verbs that take no *-ed* suffix, he generalizes this rule to all vowel-changing verbs, and only later learns the exceptions. Among some working class speakers, however, this rule may be pushing on to completion. It clearly represents an unmarking at some level of linguistic structure. If it is not stigmatized out of existence, it could point to a future direction of change of the standard dialect. Argumentation about these /t,d/'s being absent from the underlying forms is premature at this point, however, as even the most radical of the high-A speakers have at least a few of these /t,d/'s appearing in the favorable environments (before V or Q). (Note that a factor value of 1.00 does not imply 100% deletion when that factor is present in the rule environment; rather, it signifies that the factor favors deletion more than any other in its group.)

Style Shifting

We did not attempt to analyze style shifting for most of our speakers, as the analytical techniques we were using require large amounts of data. The methods developed by Labov for eliciting a range of styles in individual interviews obtain comparatively very small amounts of data on all styles other than the relatively careful general interview style. There were two speakers for whom clearly different styles were obtainable, however.

The first was Bruce H., Sr. Throughout the first half of his interview he was stiff, careful and formal, talking constantly about formal topics, especially religion. Then he went out on family business for half an hour; when he returned he was a considerably more relaxed subject, joking and laughing freely, and concentrating on different, less formal topics. The data from his interview was coded using an additional factor group for style, with the first half of the interview coded as careful and the second half as casual.

The other speaker was Alice B., whom Payne accompanied for an entire day, recording all her interactions with other people (as described in Footnote 18). For most of the day Alice was at her job at a travel agency, and the tape records her speaking to customers and ticket agents over the phone as well as conversing with her coworkers and friends. For her we distinguished two styles: All interactions with friends, family, and coworkers were coded as a casual style, and all interactions with customers and ticket agents (primarily over the telephone) were coded as a somewhat more formal, but still necessarily very friendly, style.

The results of the style factor group for these two speakers are as follows:

	Careful	Casual
Bruce H.	.74	1.00

	Business	Casual
Alice B.	.94	1.00

Both Bruce and Alice show less deletion in their more formal styles, though the difference between the two styles is clearly more substantial for Bruce than for Alice. It is interesting to note that the effect of style shifting on this rule is thus characterizable as a general upward or downward reweighting of the probabilities of deletion for all environments, rather than as involving different treatments for different environments, and this also tends to confirm the notion of a unitary variable rule as accounting for this behavior.

Conclusions

On the practical side, we feel we have developed a powerful and useful methodology for the analysis of certain types of variation with the Cedergren–Sankoff program. We have attacked the problem of data quantity directly—on the one hand we have tried to see what level of reliability can be expected from the available data, and on the other we have exploited the interview techniques developed by Payne and other members of our research group, techniques providing longer, richer interviews as well as continued access to people as more data is required.

On the substantive side, we have demonstrated that the /t,d/ deletion rule is omnipresent in a range of English speakers, and very stable and uniform with regard to its major constraints. It does show sensitivity to lectal differences in precisely those areas where a linguistic analysis would be ambivalent. Except for those differences, we suggest that it is a stable variable rule that is uniformly compelling on all speakers, and that any indications to the contrary would be due to insufficient data.

Finally, I would like to point out the similarities between the variable rule methods employed in the present study and the calculus for determining "environment weightings" developed by Bailey in *Variation and Linguistic Theory*. Bailey's method uses integral feature weights which are summed to yield weightings which impose an order on the possible environments for a given rule. Although this method is applicable to only a small number of factors and factor groups, it is basically similar in approach to the probabilistic variable rule techniques which Cedergren and Sankoff have developed and we have employed. It is very heartening to note this convergence of methodology among linguists engaged in the study of variation.

Appendix

TABLE 1.A.1
Factor Values from a Preliminary Study

	M	A	P	K	U	G	V	Q	S	$	N	F	L	Number of tokens
Philadelphians														
Jack F., 60, W Phila.	.94	1.00	.47	1.00	.47	1.00	.23	.29	.75	.38	—	1.00	.72	126
Jane R., 16, NE Phila.	.90	1.00	.90	1.00	.39	.36	.19	.15	1.00	.75	—	.85	.64	160
Johnny & Bom, 25, W Phila.	1.00	.69	.22	1.00	.87	.90	.79	.75	1.00	.91	.74	.50	1.00	133
Mary C., 41, Dorothy, NJ	.60	0.00	1.00	1.00	0.00	0.00	.48	.34	.18	.54	1.00	.10	.49	81
Non-Philadelphians														
Ed Q., 63 Ringoes, NJ	.82	1.00	.43	.63	1.00	.32	.29	.64	.77	1.00	.83	.16	.71	205
Coleman family, London	.83	.60	.60	.91	—	.43	.65	1.00	1.00	.25	.34	.19	.13	101
Mr. & Mrs. Billie W. Glasgow	1.00	.48	.34	.77	1.00	.43	.24	.48	.90	1.00	.49	.69	.42	445
Hawaiian boys	1.00	1.00	.59	1.00	.76	.76	.21	1.00	1.00	.57	—	—	.59	129
7 South & SW US whites	1.00	.66	.28	1.00	—	.65	.44	.35	.94	.93	.69	.75	.44	648
Cosmopolitan sample (all of the above)	1.00	.91	.49	1.00	.89	.60	.36	.53	1.00	.86	.65	.44	.57	1931

Acknowledgments

This research was supported by National Science Foundation Grant GS-36382X.

References

Bailey, C.-J. N. 1972. The integration of linguistic theory: Internal reconstruction and the comparative method in descriptive analysis. In R. P. Stockwell and R. K. S. Macaulay (Eds.), *Linguistic change and generative theory*. Bloomington: Indiana University Press.

Bailey, C.-J. N. 1973. The patterning of language variation. In R. W. Bailey and J. L. Robinson (Eds.), *Varieties of present-day English*. New York: Macmillan.

Bailey, C.-J. N. 1974. *Variation and linguistic theory*. Arlington, Va.: Center for Applied Linguistics.

Bailey, C.-J. N., and Shuy R. W. (Eds.). 1973. *New ways of analyzing variation in English*. Washington, D.C.: Georgetown University Press.

Bickerton, D. 1971. Inherent variability and variable rules. *Foundations of Language*, *7*, 457–492.

Bickerton, D. 1973a. Quantitative vs. dynamic paradigms: The case of Montreal 'que.' In C.-J. Bailey and R. Shuy (Eds.), *New ways of analyzing variation in English*. Washington, D.C.: Georgetown University Press.
Bickerton, D. 1973b. On the nature of a Creole continuum. *Language, 49,* 640–669.
Cedergren, H. J. 1973. On the nature of variable constraints. In C.-J. Bailey and R. Shuy (Eds.), *New ways of analyzing variation in English*. Washington, D.C.: Georgetown University Press.
Cedergren, H. J., and Sankoff, D. 1974. Variable rules: Performance as a statistical reflection of competence. *Language, 50,* 333–355.
Fasold, R. 1972. *Tense marking in Black English*. Arlington, Va.: Center for Applied Linguistics.
Griffin, P., Guy, G., and Sag, I. 1973. Variable analysis of variable data. In Language in the context of space, time and society. *University of Michigan Papers in Linguistics, 1 (2).*
Guy, G. 1975. Use and application of the Cedergren–Sankoff Variable Rule Program. In R. Fasold and R. Shuy (Eds.), *Analyzing variation in language*. Washington, D.C.: Georgetown University Press. Pp. 59–69.
Kiparsky, P. 1972. Explanation in phonology. In S. Peters (Ed.), *Goals of linguistic theory*. Englewood Cliffs, N.J.: Prentice-Hall.
Labov, W. 1966. *The social stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.
Labov, W. 1967. Some sources of reading problems for Negro speakers of non-standard English. In A. Frazier (Ed.), *New directions in elementary English*. Champaign, Ill.: National Council of Teachers of English.
Labov, W. 1969. Contraction, deletion, and inherent variability of the English copula. *Language, 45,* 715–762.
Labov, W. 1972a. The internal evolution of linguistic rules. In R. Stockwell and R. K. S. Macaulay (Eds.), *Linguistic change and generative theory*. Bloomington: Indiana University Press.
Labov, W. 1972b. Where do grammars stop? In R. Shuy (Ed.), *Georgetown monograph on languages and linguistics, 25,* 43–88.
Labov, W. 1975. The quantitative study of linguistic structure. In K.-H. Dahstedt (Ed.), *The Nordic languages and modern linguistics 2*. Stockholm: Almgrist & Wiksell International.
Labov, W., and Cohen, P. 1967. Systematic relations of standard and non-standard rules in the grammars of Negro speakers. In *Project literacy reports 8*. Ithaca, N.Y.: Cornell University.
Labov, W., Cohen, P., Robins, C., and Lewis, J. 1968. *A study of the non-standard English of Negro and Puerto Rican speakers in New York City*. Cooperative Research Report 3288. (Vols. 1 and 2.) New York: Columbia University. (Reprinted by U.S. Regional Survey, 204 North 35th Street, Philadelphia, PA 19104.)
Payne, A. 1974. The re-organization of linguistic rules: A preliminary report. *Pennsylvania working papers on linguistic change and variation II* (6).
Sankoff, G. 1972. A quantitative paradigm for the study of communicative competence. Presented to the Conference for the Ethnography of Speaking, Austin, Texas.
Sankoff, G. 1973. Above and beyond phonology in variable rules. In C.-J. Bailey and R. Shuy (Eds.), *New ways of analyzing variation in English*. Washington, D.C.: Georgetown University Press.
Summerlin, N. C. 1972. *A dialect study: Affective parameters in the deletion and substitution of consonants in the deep South*. Unpublished doctoral dissertation, Florida State University.
Wolfram, W. 1969. *A Sociolinguistic description of Detroit Negro speech*. Washington, D.C.: Center for Applied Linguistics.
Wolfram, W. 1971. *Overlapping influence in the English of second generation Puerto Rican teenagers in Harlem*. Final Report on OE Grant 3-70-0033(508). Washington, D.C.: Center for Applied Linguistics.

2

Helene Neu

RANKING OF CONSTRAINTS ON /t,d/ DELETION IN AMERICAN ENGLISH: A STATISTICAL ANALYSIS

Introduction

Phonological variation is an inherent characteristic of continuous speech. Much of this variation is systematic and can be captured by phonological rules. Work stemming from Labov (1969) has proposed the notion of variable rules to account for systematic variation. Variable rules incorporate into linguistic description the predicted relative frequency of a rule's application or rankings of constraints affecting rule application depending on linguistic and extralinguistic constraints. Examination of performance data is essential if one is to identify and rank the constraints on rule application.

Identification and ranking of significant constraints is, however, more complicated than has been indicated in many of the previous studies of variation. A frequently cited example of variation is final /t,d/ deletion in English, a rule whose constraints, it has been claimed, are fairly easily ranked. This study will examine the linguistic constraints on /t,d/ deletion in consonant + /t,d/ clusters in American English, present the results of a statistical analysis, compare the results with those of other studies, and demonstrate that the problems of ranking constraints may be more complex than has been previously believed.

The importance of statistical validation of the rankings will be emphasized, and the relationship between frequency of rule application and frequency of word occurrence will be discussed.

LOCATING LANGUAGE IN TIME AND SPACE

ISBN 0-12-432101-1