VARIABLE RULES:

PERFORMANCE AS A STATISTICAL REFLECTION OF COMPETENCE

HENRIETTA J. CEDERGREN and DAVID SANKOFF

Université du Québec à Montréal *Université de Montréal*

Speech performances are here considered as statistical samples drawn from a probabilistic language competence. This competence is modeled in conventional generative terms, except that optional rules are assigned application probabilities as functions of the structure of the input strings, possibly depending on the extralinguistic environment as well. We develop the mathematical background for these variable rules, which were introduced by Labov, and we apply the theory and methodology to examples from Spanish, French, and English. The data consist of relative frequencies of rule application for different types of input string, and they provide a check on the frequencies predicted by variable rules. This study supports the hypothesis that the various features in rule structural descriptions tend to act independently, in the statistical sense, on rule probabilities. The analytic framework for variable rules is easily and naturally extendable to take into account the effect of sociolinguistic and stylistic features on application frequencies.

1. OPTIONALITY VS. VARIABILITY AS COMPONENTS OF COMPETENCE. The systematic nature of so-called 'free' variation, and its regular interrelationships with variation according to stylistic, socio-economic, dialectal, and temporal dimensions, have stimulated a re-examination of the notion of 'optional rule' (Labov 1969, Bailey 1971, Bickerton 1971). In usual generative terms, sets of utterance tokens defined as linguistically unconditioned variants, or alternate ways of saying the same thing, have been explained in terms of the execution or non-execution of optional rules. Whereas an obligatory rule operates on all input strings that satisfy its structural description, an optional rule may or may not apply to a satisfactory input string. In these terms, no accounting is or can be made of the fact that the option is subject to regular constraints revealed through patterns of covariation with elements of the linguistic environment and with non-language factors such as age, class, and social context.

The notion of optionality fails to capture the nature of the systematic variation which exists even on the level of the grammar of a single individual. It does not permit the incorporation of relativity or covariation between the presence of certain features in the linguistic environment of a rule and the frequency of operation of the rule. The label 'optional' fails to convey any information as to how the elements of the structural description of a rule favor or constrain its operation. Rather, use of this label implies that all such information is foreign to the COMPETENCE of the native speaker.

Linguistic PERFORMANCE, on the other hand, is clearly NOT free of covariation or systematic dependence of rule execution frequency upon details of the structural description. The analysis of speech behavior has repeatedly revealed that the possibilities represented by abstract optional rules are distributed in a reproducible and well-patterned way in a given speaker and in a given speech community. Although performance may be considered only an approximate reflection of competence—because of statistical fluctuations in rule execution frequencies, and interference from physiological factors, memory limitations, and errors—it is difficult to escape the conclusion that those aspects of performance that are found to be thoroughly systematic in an individual and throughout a community are reflections of linguistic competence.

2. THE LABOV PARADIGM. In recent theory, there have been two conflicting approaches to variation and its place in linguistic competence. One approach (exemplified by Bickerton 1971; De Camp 1971; Elliott, Legum & Thompson 1969; and, to some extent, Labov 1971) attempts to account for variation on the social, dialectal, situational, and temporal levels by postulating the existence of many grammars, each occupying a position along a continuum or within some other finely distinguishable configuration of possible grammars; and on the individual level, by searching for ever more delicate distinctions to incorporate into the structural descriptions of rules, in order to eliminate or at least reduce the number of environments in which variation can occur. This approach may be characterized as an effort to down-grade or explain away variability within the traditional generative paradigm.

A radically different approach has been taken by Labov 1969, Bailey 1971, and G. Sankoff (MS), who focus on variability as a central aspect of linguistic competence. The latter two authors argue that this focus requires a different analytic paradigm from the usual generative framework; and they both point to Labov's work, particularly his article on the English copula, as a beginning of the 'paradigm change', in the sense of Kuhn 1970. Both document this change in terms of altered goals, data sources, analytic techniques, and theory content.

The key to the new paradigm lies in Labov's proposal to incorporate systematic variation into linguistic description and theory by extending the concept of a rule of grammar to that of a VARIABLE RULE, where the predicted relative frequency of a rule's operation is, in effect, an integral part of its structural description. This single idea, which is essentially a proposal about individual linguistic competence, also has wide ramifications for stylistics, sociolinguistics, dialectology, and diachronic linguistics. This stems from the fact that, once accepted and incorporated into description, variability can be made a function not only of the presence or absence of linguistic elements, but also can be constrained by extralinguistic factors, all within the same notational and theoretical framework.

3. SOCIAL AND INDIVIDUAL GRAMMARS. Linguistic theory has traditionally separated language, as a homeostatic organism, from the erratic performance of individual speakers. The discovery of invariant structure—the linguistic knowledge of the ideal speaker–hearer of a homogeneous speech community—has been the primary interest.

The fundamental problem in linguistics is the description of this knowledge in the form of grammar. The grammars written by linguists are assumed to be those of speech communities, although the analytic assumption of homogeneous invariant structure has led to a weak statement of goal, viz. the description of the ideal native speaker's knowledge. The generalizing capabilities of the grammar depend on the size of the homogeneous speech community. In practice, this is often limited, in effect, to ONE member.

A model of language as an orderly heterogeneous structure, as proposed by Weinreich, Labov & Herzog 1968, would eliminate the search for ideal speakers and make speech communities directly accessible to the linguist. Variation would be placed against the background of the community. Indeed, we shall show later how certain phonological and morphological entities, whose manifestations are extremely heterogeneous on the community level, can each be accounted for by a single rule, which both describes the range of variation present in the community and accurately predicts the behavior of each individual.

The variable rules developed by Labov should, like other rules of generative grammar, be interpreted as part of individual competence. The numerical quantities associated with the features in the environment of a rule are indications of the relative weight which they contribute to the applicability of the rule, rather than the existence of discrete probabilities in the head of the speaker. Their precise values, as in any behavioral model, are not critical; they represent analytical abstractions of tendencies which may vary somewhat from day to day or from speaker to speaker. Nevertheless, these tendencies are quite real; long series of utterances which display SIGNIFICANTLY discrepant frequencies are unlikely to be produced by a native speaker, and are easily detected as unnatural (ungrammatical?) by native hearers. What discrepancies are significant, what level of difference in probabilities between two speakers or two speech varieties can be perceived by native hearers, how finely speakers can control or vary their rates of rule application—these are all empirical problems which may have different answers in different contexts. Thus, although we will write frequencies and probabilities to an accuracy of two or three decimals, we certainly make no assumption that differences of 0.01 or even 0.1 are subjectively significant.[1]

Labov's efforts to draw attention to the probabilistic aspects of grammar have met with resistance from many students of language. This attitude, best expressed by Bickerton 1971, has two components, one a methodological claim and the other a position on the psychology of language. The first is a belief that all linguistic variability eventually can be deterministically accounted for, through careful enough attention to linguistic and social-interactional conditioning; the second is a belief that language processing capabilities, including those related to acquisition, could not possibly include a random or non-deterministic component. G. Sankoff (MS) has shown in some detail that, while the first claim is of course logically incapable of being disproved by data, it is certainly not warranted by any of the relevant existing data or analyses. The second claim ignores the extensive psychological literature documenting probabilistic aspects of mental processes, such as 'probability matching'.[2]

4. PROBABILISTIC MODELS OF VARIABILITY. In an optional rule for rewriting an element of a string, many aspects of the linguistic environment of the element may be observed to affect the probability of application of the rule: the detailed nature

[1] The problem of the significance of differences is universally encountered with continuous scales of measurement, especially in the behavioral sciences; but this is no indication that such scales are not applicable or meaningful.

[2] See, e.g., Estes 1964 and Lee 1971. A linguistically relevant example is Tucker's 1967 experiments on the relation of gender to word-ending in French.

of the preceding or following phonological segment, the grammatical character of the morphological unit containing the element, its position relative to word boundaries, etc. If each of these aspects may be categorized in several ways, the total number of analytically distinct environments may become immense.[3] Measuring and calculating rule application probabilities for each of these environments separately would result in an unwieldy and unenlightening mass of numbers, difficult to incorporate into generative notation.

Linguistic method traditionally starts from a study of individual cases, and attempts to assemble them into rules of maximum generality. Labov has extended this principle to the quantitative analysis of rule application and discovered the highly significant generalization that the presence of a given feature or subcategory tends to affect rule frequency in a probabilistically uniform way in all the environments containing it. Thus a preceding vowel favors copula contraction, regardless of the nature of the grammatical constituent following the copula; the presence of a following verb also favors contraction, whether the preceding segment is a vowel or consonant. This means that, instead of having to account for all the different possible environments in an unconnected way, which would be counter to usual linguistic practice, it suffices to calculate the effect or contribution of each environmental feature (these are generally much less numerous than environments),[4] and to know how to combine these effects to calculate probabilities for the specific environments. Labov's discovery that all environments for a given rule are governed by a fixed set of feature effects, combining in a highly predictable way, is a novel and strong justification for the usual practice of notationally collapsing similar disjunctively ordered rules, operating in different environments, into a single general rule.

The statement that a given feature tends to have a fixed effect independent of the other aspects of the environment may be mathematically formulated in a number of different ways. In choosing among them, we must consider three basic criteria. First, the formulation must predict rule frequencies that jibe with the observed data. Second, it must be as universally applicable as possible; a different type of mathematical structure for each variable rule is to be avoided. Third, the formulation should have some reasonable interpretation as a statement about linguistic competence.

Perhaps the most straightforward formulation is the 'additive' model, asserting that the rule probability p in a given environment is simply the sum of a number of quantities, one for each relevant feature in the environment:

$$(1)\quad p = p_0 + \alpha_i + \alpha_j + \ldots$$

where α_i is a fixed number which enters into the formula if and only if feature i is present in the environment, and p_0 is an 'input' common to all environments. Note that the effect of a given feature depends only on its presence and not on the other aspects of the environment. This is the type of model on which the statistical

[3] E.g., if there are x types of preceding segment, y types of following segment, z types of morphological unit, etc., this usually implies about $x \times y \times z \times \ldots$ different types of environment.

[4] I.e., when there are about $x \times y \times z \times \ldots$ environments, there are only $x + y + z + \ldots$ features.

technique of 'analysis of variance' is based. An additive model was successfully employed by Labov 1969 in his study of contraction and deletion of the copula in English.

How, in general, does the additive model perform in terms of the three criteria we have set forth? First, it fits the data reasonably well for the examples in which it has been used. Second, it is applicable in a wide class of rules. But there is another large class of rules to which the additive model is not applicable in its unmodified form. Since probabilities in general, and variable rule application probabilities in particular, are numbers between zero and one, a general model for variable rules should not be capable of predicting application 'probabilities' outside the interval between zero and one. For some rules, however—especially when application frequencies are very different in different environments, or when there are a large number of different environments—an ordinary additive model, fitted by conventional statistical techniques, will tend to predict such meaningless values. There are several ways of modifying the model to avoid this; Labov's suggestion of 'strong geometric ordering' was an assumption about the relative sizes of the various feature effects, formulated to preclude sums outside the zero-to-one interval. Unfortunately, this assumption is borne out by the data only in some cases. Another way to modify the additive model is to adopt the 'truncating' convention: when the sum of the effects is greater than one, the rule probability is exactly one; and when the sum is less than zero, the probability is zero.

The third criterion for judging the mathematical model deals with its linguistic import, and here the additive model has been criticized on the grounds that it posits a numerical computation facility as part of competence. We would not agree with this criticism, which is essentially of the same genre as the empirically unjustified 'psychological reality' argument discussed in §3 above; but we would agree that a formulation of variable rules should make as much use as possible of the conventional generative conceptual framework for rules.

One of the main objectives of this article is to introduce another model, the multiplicative model, for analysing variable rules. This, though similar to the additive model and retaining many of its useful properties, seems superior to it by all three criteria. In the multiplicative model, each factor is associated with a fixed effect; but instead of adding the effects, we multiply them together to arrive at the value pertinent to a given environment. One important complication is introduced along with this type of model: viz. the problem of deciding, for a given rule, whether it is the application probabilities or non-application probabilities which obey the multiplicative law. This question will be discussed in detail in the next section; here we will illustrate with multiplicative non-application probabilities. If we retain the symbol p for application probabilities, then $1-p$ is the probability that the rule does not apply, and the model is summarized by

(2) $(1-p) = (1-p_0) \times (1-p_i) \times (1-p_j) \times \ldots$

where p_0 is, as before, an input probability common to all environments, and p_i can be considered to the probability contribution of feature i, so that the factor $(1-p_i)$ is present or absent from the formula depending on whether or not feature i is present or absent from the environment. The p_i are all between zero and one.

How does the multiplicative model compare to the additive model by each of our three criteria? First, for the twenty-odd rules on which we have data, the multiplicative model almost always fits these data as well as or better than an additive model. (In many cases, it should be mentioned, there is little reason to choose among an additive model, a multiplicative application probabilities model, and a multiplicative non-application probabilities model, when data-fitting is the sole criterion.) Second, with respect to universality, a given multiplicative model is always directly applicable as is, even where additive models must be truncated or otherwise complicated, since the former predicts only bona-fide probabilities (i.e. between zero and one). Third, multiplicative models lend themselves to a very simple interpretation as to the nature of the probabilistic component of linguistic competence. In effect this links Labov's discovery of the independence of feature contribution to the notion of independence in the probabilistic sense. Furthermore, the idea of combining variable constraints becomes a natural and mild extension to the concepts of conjunctive and disjunctive ordering, depending on whether it is application probabilities or non-application probabilities which are multiplicative. These interpretations will be discussed in the next two sections.

5. INDEPENDENCE VS. INTERACTION OF ENVIRONMENTAL FEATURES. In probability theory, for a number of events to be mutually INDEPENDENT, it is required that the probability that ALL of them occur be exactly equal to the product of all their individual occurrence probabilities. If, on a certain day, the probability of rain is $Pr\{\text{rain}\} = \frac{1}{3}$, and the probability of high winds is $Pr\{\text{wind}\} = \frac{1}{2}$, then the two meteorological events are independent if and only if

$$\begin{aligned} Pr\{\text{both rain AND wind}\} &= Pr\{\text{rain}\} \times Pr\{\text{wind}\} \\ &= \tfrac{1}{3} \times \tfrac{1}{2} \\ &= \tfrac{1}{6}. \end{aligned}$$

From this we derive the notion of INDEPENDENT EXPERIMENTS. A number of experiments are said to be independent if the possible outcomes from each individual experiment are all events which are independent of the outcomes of all the other experiments. E.g., the weather today and the weather tomorrow can be considered independent experiments if each of the events 'rain today', 'sun today', 'wind today' etc. is probabilistically independent of each of the events 'rain tomorrow', 'sun tomorrow', 'wind tomorrow' etc.

In speaking of independent experiments, there is usually an explicit or implicit temporal ordering of these experiments; but no time variable appears in the definition, nor is one implicit. In fact, we shall concentrate on, and develop the mathematics for, a special case where all the experiments take place concurrently.

Suppose any one of a number of environmental features can be thought of as being able to cause, IN A PROBABILISTIC SENSE, the successful application of a rule. I.e., the presence of feature i will, with probability p_i, cause the rule to apply regardless of which other features are also present. Suppose further that, in a specific environment, the effects of all the features present constitute independent experiments, and that the rule applies if AT LEAST ONE of the features so causes it. (This disjunctive ordering of experiments is a generalization of the ordinary disjunctive ordering of constraints, in which case each of the $p_i = 1$.) The overall probability p of rule application is then

$$(3)\ 1 - (1-p_i) \times (1-p_j) \times \ldots$$

where there is one factor $(1-p_k)$ for each feature k present in the environment.[5] This is just the multiplicative non-application probabilities model (2), except that we still must introduce the input probability p_0. This will be done in §6, below.

To derive the multiplicative application probabilities model, we assume that, instead of each feature i being able to cause rule application, the feature PERMITS the rule application with probability p_i. Then the assumption of independence, together with the condition that ALL the features in the environment must permit the rule for it to apply, leads to the formula

$$(4)\ p = p_i \times p_j \times \ldots$$

where there is one feature p_k for each feature k present in the environment. (This is a generalization of conjunctive ordering of constraints. Each feature i in an ordinary conjunctively ordered condition for application can be considered to have $p_i = 1$, and each feature k not compatible with the condition has $p_k = 0$.) As with the non-application probabilities model, it will be advantageous to introduce an input probability p_0.

Note that, in both multiplicative models, feature i's contribution p_i to rule application probabilities is itself interpretable as a probability. This is not the case with additive models, where the coefficients are not automatically interpretable as probabilities.

In any empirical context it is not necessarily true that factors operate independently on the probabilities of events. Indeed, there are any number of examples where factors interact in different ways. In testing the effects of two drugs, e.g., we may find they act synergistically in curing a disease: $Pr\{$no cure by drug A combined with drug B$\}$ may be much less than $Pr\{$no cure by drug A alone$\} \times Pr\{$no cure by drug B alone$\}$. Or they might act antagonistically; i.e., the probability of no cure by the combination of drugs may be GREATER than the product of the individual probabilities. Nevertheless, unless there is some theoretical reason to suspect one of these kinds of interaction, the assumption of independence seems the most reasonable and non-committal; at least, independence is the null hypothesis suitable for a wide range of empirical situations. Thus, in investigating the relationship between the different components of the linguistic environment, or between the linguistic and the sociolinguistic environments, as factors influencing the application of a rule, we will not know a-priori whether or not they act independently; rather, we will assume independence, and then see how well the resultant models fit the data.

Note that our models in no way preclude the concurrent consideration of purely linguistic, sociolinguistic, stylistic, and other types of environmental features. Indeed, by easing Labov's restrictions on the ordering of constraints, we are free

[5] The probability p is the sum of the probabilities of all the experiment outcome configurations containing at least one successful experiment; this includes every configuration except the single one in which the rule successfully applies in NONE of the experiments. Thus p equals 1 minus the probability of this latter configuration which, by independence, is just $(1-p_i) \times (1-p_j) \times \ldots$

to experiment with as many features and types of features as might be relevant to a given rule, to combine them in a mathematically and linguistically significant way, and to compare the effects of the same feature across several different rules.

6. NOTATION AND RULE STRUCTURE. The notation we shall adopt represents a convergence in usage among writers of variable rules.[6] Where an optional rule is traditionally written as

$$\text{(5)}\quad X \rightarrow (Y) \,/\, \left\{\begin{matrix}[\text{fea A}]\\ [\text{fea B}]\\ \vdots\end{matrix}\right\} \left\{\begin{matrix}\underline{\qquad}\\ [\text{fea I}]\\ [\text{fea J}]\\ \vdots\end{matrix}\right\} \left\{\begin{matrix}[\text{fea P}]\\ [\text{fea Q}]\\ \vdots\end{matrix}\right\} [\text{fea Z}]$$

the corresponding variable rule will be written

$$\text{(6)}\quad X \longrightarrow \langle Y\rangle \,/\, \left\langle\begin{matrix}[\text{fea A}]\\ [\text{fea B}]\\ \vdots\end{matrix}\right\rangle \left\langle\begin{matrix}\underline{\qquad}\\ [\text{fea I}]\\ [\text{fea J}]\\ \vdots\end{matrix}\right\rangle \left\langle\begin{matrix}[\text{fea P}]\\ [\text{fea Q}]\\ \vdots\end{matrix}\right\rangle [\text{fea Z}]$$

where each pair of angled brackets in the structural description contains a list of the features, feature bundles, subcategories etc., which may be pertinent to the relevant position in the string. As usual, any feature not in angled brackets, e.g. [fea Z], denotes a minimally required feature: it must be present in the environment for the rule to apply. In the simplest cases, each list in angled brackets is mutually exclusive and exhaustive for that position; i.e., exactly one item in the list occurs in each environment of X. E.g., if both [fea R] and [fea S] must be either plus or minus in the preceding segment,[7] this might be written as either

$$\text{(7a)}\ \left\langle\begin{matrix}\begin{bmatrix}+R\\+S\end{bmatrix}\\ \begin{bmatrix}+R\\-S\end{bmatrix}\\ \begin{bmatrix}-R\\+S\end{bmatrix}\\ \begin{bmatrix}-R\\-S\end{bmatrix}\end{matrix}\right\rangle \underline{\qquad} \qquad \text{or} \qquad \text{(7b)}\ \begin{matrix}\left\langle\begin{matrix}[+R]\\ [-R]\end{matrix}\right\rangle\\ \left\langle\begin{matrix}[+S]\\ [-S]\end{matrix}\right\rangle\end{matrix} \underline{\qquad}$$

or as some less redundant abbreviation of one of these two, such as

$$\text{(8a)}\ \left\langle\begin{matrix}\begin{bmatrix}+R\\+S\end{bmatrix}\\ [+R]\\ [+S]\end{matrix}\right\rangle \underline{\qquad} \qquad \text{or} \qquad \text{(8b)}\ \begin{matrix}\langle[+R]\rangle\\ \langle[+S]\rangle\end{matrix} \underline{\qquad}$$

The point is that these notations all satisfy the mutually exclusive and exhaustive conditions, 7a and 8a by listing all possible combinations of the two features, 7b and 8b by considering them as occupying two distinct positions in the structural description, i.e. as two distinct variable constraints.

[6] The details of this notation were worked out in consultation with Labov (see Labov 1973a, b).

[7] Our notation is not specific to binary features. Thus [fea A] and [fea B] in rule 6 could be [+R] and [−R] respectively; but in another rule they could be just two possibilities among six or seven different feature bundles or syntactic categories.

Whenever the exhaustivity condition does not appear to be met, it can always be satisfied either by the implicit addition of a residual category, as in 8a–b, or by the implicit listing of the null element $\emptyset$. The requirement of mutual exclusivity of the features in a list may also be relaxed somewhat. Some of the features may be subcategories of others in the same list, as long as no one feature is a subcategory of two or more disjoint (i.e. exclusive) features, and as long as each environment of the variable contains exactly one feature from the list which is not further subcategorized plus all the features of which it is a subcategory. E.g., in a list containing A, B, C, D, E, F, G, and H, features A and B might be subcategories of E, C and D subcategories of F, and E and G subcategories of H, as in Figure 1.

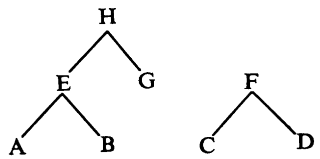

FIGURE 1

Then each environment must contain exactly one of A, B, C, D, or G; and if an environment contains B, for example, it must also contain E and H. By relaxing the exclusivity requirement in this way, we can handle variable constraints which include hierarchically arranged features, such as phonological distinctive features.

Although all possible linguistic environments must be accounted for in the variable rule notation, it is not necessary that all the combinations of features, one from each list, specify linguistically possible environments. It is normal to avoid redundancy in rule notation; and since impossible environments are precisely those which cannot arise through previous rules acting on underlying forms, it seems undesirable to invent any new conventions to distinguish possible from impossible environments again in the variable rule. Furthermore, the impossible environments will usually be distinguished in data presentations as empty cells.

Associated with each feature or variable constraint is a probability, i.e. a number between zero and one. It is most convenient, and least disruptive of conventional notation, to summarize these probabilities in a small table separate from the rule proper. In addition, this table contains the input probability p_0. In the multiplicative non-application probabilities model, this input is the probability that the rule will apply in the least favorable environment for the rule, and often $p_0 = 0$. In the application probabilities model, p_0 can be considered as the rule probability in the most favorable environment. It is important to incorporate an input parameter into the model whenever there are two or more lists of environmental features in the structural description of a rule. Otherwise all the probabilities associated with one list of features could be arbitrarily increased, within a certain range, as long as all the probabilities for another of the lists were correspondingly decreased, without changing rule probabilities. This would make feature effect comparison across lists and across rules relatively meaningless. Once the input probability has been properly defined, however, this arbitrariness is removed from the model, and it becomes meaningful to make such cross-list comparisons. Another consequence of these conventions is that in every list of features, feature bundles, or subcategories, one of them (often the residual category) can be singled out as having no effect on rule probability, i.e. $p_i = 0$ (non-application model) or $p_i = 1$ (application model).

E.g., consider the variable constraints of 7b, where $\langle[+\mathrm{R}], [-\mathrm{R}]\rangle\underline{\hspace{2em}}$ and $\langle[+\mathrm{S}], [-\mathrm{S}]\rangle\underline{\hspace{2em}}$ represent two different variable constraints, belonging to a multiplicative non-application probabilities model, where the environmental frequencies are: for $[+\mathrm{R}, +\mathrm{S}]\underline{\hspace{1em}}$, 0.60; for $[+\mathrm{R}, -\mathrm{S}]\underline{\hspace{1em}}$, 0.72; for $[-\mathrm{R}, +\mathrm{S}]\underline{\hspace{1em}}$, 0.20; and for $[-\mathrm{R}, -\mathrm{S}]\underline{\hspace{1em}}$, 0.44. Then, lacking the notion of input probability, we could not distinguish between $p_{+\mathrm{R}} = 0.6$, $p_{-\mathrm{R}} = 0.2$, $p_{+\mathrm{S}} = 0$, $p_{-\mathrm{S}} = 0.3$ and an alternative set of values $p_{+\mathrm{R}} = 0.55, p_{-\mathrm{R}} = 0.1, p_{+\mathrm{S}} = 0.11, p_{-\mathrm{S}} = 0.38$— or between either of these and any other of a large class of models, since they all predict exactly the same environmental frequencies. But if p_0 must be the predicted frequency in the least favorable environment, then all the probabilities are uniquely determined at $p_0 = 0.2, p_{+\mathrm{R}} = 0.5, p_{-\mathrm{R}} = 0, p_{+\mathrm{S}} = 0, p_{-\mathrm{S}} = 0.3$.

The input parameter also serves the purpose of containing 'unexplained variability', in the sense that, as new environmental factors are discovered to play a role in influencing the rule probability, they may be incorporated into the rule as a new series of features; e.g., rule 6 might become

$$\text{(9) X} \longrightarrow \langle \text{Y} \rangle \;/\; \left\langle \begin{matrix} [\text{fea A}] \\ [\text{fea B}] \\ \vdots \end{matrix} \right\rangle \left\langle \begin{matrix} \overline{[\text{fea I}]} \\ [\text{fea J}] \\ \vdots \end{matrix} \right\rangle \left\langle \begin{matrix} [\text{fea P}] \\ [\text{fea Q}] \\ \vdots \end{matrix} \right\rangle [\text{fea Z}] \left\langle \begin{matrix} [\text{fea S}] \\ [\text{fea T}] \\ \vdots \end{matrix} \right\rangle$$

with the appropriate adjustment of p_0 (and/or some other parameters in the original formulation).

In general, where the multiplicative formulation predicts rule probabilities greatly at variance with observed count proportions of rule applications per number of eligible environments, we reject the hypothesis of independence and examine the environmental factors more carefully, to see which ones interact and to determine the nature and linguistic significance of this interaction (cf. Cedergren 1973a).

7. PROBABILITIES VS. STATISTICS. In the preceding, we have restricted our terminology to probabilities and predicted frequencies, and have not used such notions as statistics, actual observed frequencies, or proportions. This is in keeping with an important distinction between probabilities and frequencies, one which has unfortunately been overlooked in the literature on variable rules. Statistics, frequencies, estimates, and the like are all random variables, i.e., they cannot be predicted with 100% accuracy; and they vary somewhat between performances of an identical experiment. Probabilities, on the other hand, are fixed numbers, and are not subject to random variation. Sometimes probabilities are known, such as the probabilities of certain poker hands, or the probabilities imposed by an experimenter in a simulation experiment (see Klein 1965). At other times probabilities are not known, and the problem is to estimate them using observed frequencies from experiments. In any case, probabilities are assumed to be well-defined, fixed numbers inherent in the underlying model which generates the observed frequencies. Thus, if one did not have the mathematical skill to calculate the probability

of a 'full house' in poker, one could estimate this probability by counting the proportion of full houses in a large number of randomly dealt hands. In the same way we wish to estimate rule probabilities by observing rule frequencies. Frequencies are clearly part of performance; but we use them to estimate probabilities, which are inherent in the ability to generate the observed behavior. It is our contention that these probabilities are properly part of competence.

There are many ways of estimating probabilities from frequencies; they can result in slightly different estimates, and we do not argue for any particular one. We use the well-known method of maximum likelihood, but we will only sketch the mathematical details lightly.

8. METHOD OF COMPUTATION. Once all the information of generative relevance is incorporated into a variable rule, the values of the feature probabilities remain to be estimated. The data on which this estimation is based consist of observations of the number of times a rule has applied, out of some total number of cases examined for a given environment. These data should be available for all possible environments in which the rule can apply. They may be arranged in various formats, as in Table 1, which is particularly suitable for computer input, or else as in Tables

FEATURES PRESENT	NUMBER OF RULE APPLICATIONS	TOTAL CASES EXAMINED	NUMBER OF NON-APPLICATIONS
A I P	45	250	205
A I Q	16	190	174
A J P	124	354	230
⋮	⋮	⋮	⋮

TABLE 1

2, 5, 8, and 9 below, which all contain non-application frequencies. As we have mentioned, a number of estimation techniques are feasible for calculating the probabilistic aspects of the rule, on the basis of frequency data, including heuristic trial-and-error methods if the number of feature configurations possible is not too great. The method of maximum likelihood finds those values of the coefficients which maximize the probability that the observed data would be produced by the model. These values can be calculated on a computer using iterative maximization methods. A program which does this, written in FORTRAN and amply documented, is available from the authors. It calculates and compares maximum likelihood solutions for an additive model (with truncation if necessary), and both application and non-application probabilities multiplicative models for any variable rule data not involving more than 20 features or 100 environments. It is easily modified to do more complicated rules.

Instead of incorporating numbers into rule notation, we accompany each rule with a table, which also includes values of stylistic and/or sociolinguistic parameters. These latter enter into probability calculations according to the same multiplicative-independence model, but we refrain from introducing them in the rule notation itself.

In the next three sections, we proceed to apply the theory and methodology we have been developing to three sets of data from different sources. These illustrate both multiplicative non-application probabilities and multiplicative application probabilities models.

9. *R*-SPIRANTIZATION IN PANAMANIAN SPANISH. Syllable-final *r* in Panamanian Spanish is realized variably as an alveolar flap [ɾ], as an apical spirant [ɾ̌], as a voiceless pharyngeal spirant [h], or as ∅, complete absence. This variation is found at all levels of Panamanian society (Ricord 1971), and in fact is widely distributed in Spanish dialects (Navarro-Tomás 1963, Zamora 1967, Canfield 1962). The range and conditioning factors of the variation have not been made explicit in the literature. Our discussion will be limited to a postulated rule of the form

$$(10)\quad \underset{\begin{bmatrix}-\text{lat}\\+\text{cns}\\+\text{voc}\end{bmatrix}}{\text{ɾ}} \rightarrow \underset{\begin{bmatrix}-\text{voc}\\-\text{ant}\end{bmatrix}}{\check{\text{ɾ}}}$$

as the first step in the derivation of the variant forms.

The information presented is based on the analysis of the speech of a sample of 79 native Panamanians resident in Panama City (Cedergren 1973b). This sample is stratified by socio-economic level, as indicated in Table 2, as well as by sex, age, and urban or rural origin. The speakers were interviewed on a variety of topics, such as childhood games, carnival customs etc. In examining the variable distribution of *r*, the following linguistic factors are discovered to be relevant:

(a) whether *r* is in internal or final position in the phonological word:

(11) *#sabor#* 'flavor' (final)
(12) *#arco#* 'arch' (internal)

(b) whether or not *r* is a formative;

(13) *#canta+r#* 'to sing' (infinitive)
(14) *#mar#* 'sea' (monomorphemic)

(c) the phonetic nature of the following segment;

(15)	*#corte#* 'court',	*#cantar##toda#* 'to sing all' (obstruent)
(16)	*#burla#* 'teases',	*#pintar##locos#* 'to paint fools' (lateral)
(17)	*#arma#* 'weapon',	*#comer##macarrones#* 'to eat macaroni' (nasal)
(18)		*#contar##ojos#* 'to count eyes' (vowel)
(19)		*¿A ver?* 'What do you want?' (pause)

No combination of factors seems to result in a categorical rule (i.e. an environment in which the rule always applies, or one in which it never applies). Thus the form of the variable rule for *r*-spirantization is postulated as follows:[8]

$$(20)\quad \underset{}{\overset{\text{ɾ}}{\begin{bmatrix}-\text{lat}\\+\text{cns}\\+\text{voc}\end{bmatrix}}} \longrightarrow \left\langle \overset{\check{\text{ɾ}}}{\begin{bmatrix}-\text{voc}\\-\text{ant}\end{bmatrix}} \right\rangle / \; [-\text{cns}]\,\langle+\rangle \;\_\_\_\; \langle\#\#\rangle \left\langle \begin{matrix}[+\text{lat}]\\ [+\text{obs}]\\ [+\text{nas}]\\ [-\text{cns}]\\ [-\text{seg}]\end{matrix} \right\rangle$$

[8] Note that, e.g., [−cns]__[−cns] or [−cns]#__[+lat] are among the linguistically impossible environments for syllable-final *r* (cf. the discussion in §6, above). Thus, in Table 2, there are no data for internal *r* preceding a vowel or pause, and none for internal inflected *r*.

Table 2 shows the observed distribution of [ɾ] for the total number of cases examined for each linguistic environment, where the data are aggregated within each socio-economic group.

It remains to estimate the feature effects in the variable rule which yield a best fit to this data. Our procedure gives the values in Table 3 for the non-application probabilities multiplicative model, which is the most consistent with these data.

	__[+obs]	__[+lat]	__[+nas]	__[−cns]	__[−seg]
INTERNAL	$\frac{133}{153}$ (0.87)	$\frac{4}{9}$ (0.44)	$\frac{29}{35}$ (0.83)		
FINAL					
monomorphemic	$\frac{25}{27}$ (0.93)	$\frac{6}{17}$ (0.35)	$\frac{7}{8}$ (0.88)	$\frac{24}{27}$ (0.89)	$\frac{11}{12}$ (0.92)
infinitive	$\frac{39}{58}$ (0.67)	$\frac{19}{30}$ (0.63)	$\frac{8}{11}$ (0.73)	$\frac{31}{50}$ (0.62)	$\frac{10}{17}$ (0.59)

Group I (high socio-economic status, 6 speakers)

INTERNAL	$\frac{533}{640}$ (0.83)	$\frac{20}{41}$ (0.49)	$\frac{112}{125}$ (0.90)		
FINAL					
monomorphemic	$\frac{79}{108}$ (0.73)	$\frac{9}{50}$ (0.18)	$\frac{13}{24}$ (0.54)	$\frac{93}{104}$ (0.89)	$\frac{40}{50}$ (0.80)
infinitive	$\frac{103}{195}$ (0.53)	$\frac{41}{127}$ (0.32)	$\frac{28}{61}$ (0.46)	$\frac{137}{220}$ (0.62)	$\frac{42}{71}$ (0.59)

Group II (19 speakers)

INTERNAL	$\frac{384}{512}$ (0.75)	$\frac{16}{29}$ (0.55)	$\frac{68}{85}$ (0.80)		
FINAL					
monomorphemic	$\frac{60}{89}$ (0.67)	$\frac{10}{37}$ (0.27)	$\frac{15}{19}$ (0.79)	$\frac{73}{93}$ (0.79)	$\frac{28}{42}$ (0.67)
infinitive	$\frac{60}{154}$ (0.39)	$\frac{25}{90}$ (0.28)	$\frac{33}{64}$ (0.52)	$\frac{85}{152}$ (0.56)	$\frac{30}{68}$ (0.44)

Group III (19 speakers)

INTERNAL	$\frac{509}{667}$ (0.76)	$\frac{43}{70}$ (0.62)	$\frac{129}{152}$ (0.85)		
FINAL					
monomorphemic	$\frac{60}{87}$ (0.69)	$\frac{13}{54}$ (0.24)	$\frac{15}{22}$ (0.68)	$\frac{86}{93}$ (0.93)	$\frac{30}{40}$ (0.75)
infinitive	$\frac{67}{177}$ (0.38)	$\frac{37}{119}$ (0.31)	$\frac{39}{89}$ (0.44)	$\frac{123}{221}$ (0.56)	$\frac{40}{100}$ (0.40)

Group IV (low socio-economic status, 35 speakers)

TABLE 2. Observed frequency of [ɾ] per total cases of *r*, and relative frequency, by position in word, morphemic status, following phonological environment, and socio-economic level of speaker.

Input probability $p_0 = 0$.

Morphemic status:	Infinitive		Monomorphemic		
Effect:	0.32		0		
Position relative to boundary:	_##		Internal		
Effect:	0.10		0		
Segmental environment	_[+obs]	_[+lat]	_[+nas]	_[−cns]	_[−seg]
Effect:	0.16	0.50	0.11	0	0.12
Socio-economic level:	I	II	III	IV	
Effect:	0	0.013	0.115	0.083	

TABLE 3. Contribution of each factor influencing *r*-spirantization, according to a multiplicative non-application probabilities model.

The extent to which our variable rule 20 captures the systematicity of the variation evident in the data is clear from Table 4. Here the probability of rule non-application in each environment has been calculated by the appropriate formula of type 2, and then multiplied by the number of times this environment occurs. The resulting predicted frequency of [ɾ] is compared with the observed frequency in this environment.

		_[+obs]	_[+lat]	_[+nas]	_[−cns]	_[−seg]
GROUP I:	INTERNAL	128.9	4.5	31.1		
		(133)	(4)	(29)		
	FINAL					
	monomorphemic	20.5	7.6	6.4	24.3	9.5
		(25)	(6)	(7)	(24)	(11)
	infinitive	29.8	9.1	6.0	30.5	9.1
		(39)	(19)	(8)	(31)	(10)
GROUP II:	INTERNAL		532.3	20.2	109.6	
		(533)	(20)	(112)		
	FINAL					
	monomorphemic	80.8	22.1	18.9	92.3	39.0
		(79)	(9)	(13)	(93)	(40)
	infinitive	98.8	38.0	32.6	132.3	37.5
		(103)	(41)	(28)	(137)	(42)
GROUP III:	INTERNAL		381.6	12.8	66.8	
		(384)	(16)	(68)		
	FINAL					
	monomorphemic	59.7	14.7	13.4	74.0	29.4
		(60)	(10)	(15)	(73)	(28)
	infinitive	69.9	24.2	30.6	81.9	32.2
		(60)	(25)	(33)	(85)	(30)
GROUP IV:	INTERNAL		515.5	32.0	123.8	
		(509)	(43)	(129)		
	FINAL					
	monomorphemic	60.5	22.2	16.1	76.7	29.0
		(60)	(13)	(15)	(86)	(30)
	infinitive	83.3	33.1	44.2	123.5	49.1
		(67)	(37)	(39)	(123)	(40)

TABLE 4. Frequency of [ɾ] predicted by variable rule 20 compared with observed frequencies (in brackets). See Table 2 for total numbers.

From Table 3, we deduce that the single environmental factor most conducive to rule application is the feature [+lateral] following the variable. Another feature which encourages spirantization of the variable is [+infinitive].

The effect of socio-economic level of speaker on rule probability is not as great as the linguistic conditioning, but it is not negligible. Similar analyses were carried out incorporating stylistic variation, and using age and residence history as non-linguistic factors. Indeed, all of these and any number of further factors could be included in a single analysis. However, it is a realistic mathematical property of multiplicative models, inherent in formulae 2 and 4, that only relatively few independent factors, linguistic or non-linguistic, can have any appreciable effect on a single rule. This is one reason why we need no ASSUMPTION such as Labov's 'strong geometric ordering' about the relative contributions of various factors. Another important and valuable property is that, with any reasonably distributed sample of speakers, the addition of social or further independent linguistic factors to a rule has very little effect on the coefficients of phonological or grammatical features previously calculated.

To illustrate that the rule constructed from data on a sample of the speech community is relevant to the description of the grammars of individual speakers, we carried out the following experiment. For each of the 79 speakers, we set the input probability at 0, 0.013, 0.115, or 0.083, depending on the socio-economic group. We then calculated the rule probabilities for each linguistic environment, and multiplied the result by the number of times the environment occurred in that speaker's interview. We then compared the predicted occurrences of rule applications to the observed occurrences. For almost all speakers, the fit was quite good. In terms of chi-square, for 80% of the informants, the discrepancy was not significant at the 5% level—this despite the fact that chi-square is too stringent a test in cases like ours where measurement errors and marginal cases tend to exaggerate the apparent sampling fluctuation, and despite our using aggregate values for p_0, instead of adjusting for individual variation within socio-economic groups. This experiment refutes any suggestion that, for this rule, the calculation of application probabilities from aggregate data has no relevance to individual grammars.

10. DELETION OF COMPLEMENTIZER *que* IN MONTREAL FRENCH. In Montreal French (MF), the complementizer *que* is often deleted in various contexts. There is no difference in meaning between the following:

(21) *Au début je pense ça a été plutôt un snobisme* 'At the beginning I think it was rather a fad.'

(22) *Au début je pense que ça a été plutôt un snobisme.*

The distribution of such sentences (i.e. with and without *que*) is subject to both linguistic and social constraints, and these we investigated in terms of a rule of the form

(23) *que* $\rightarrow \emptyset$.

We used data from recorded interviews of 16 Montrealers, categorized according to occupational class and sex as indicated in Table 5. These represent part of a larger sample of 120 French-speaking Montrealers interviewed for a study of variation in MF (G. Sankoff & Cedergren 1972, D. Sankoff & G. Sankoff 1973).

	WORKING-CLASS MEN			WORKING-CLASS WOMEN		
	__[+sib]	$\_\begin{bmatrix}+\text{cns}\\-\text{sib}\end{bmatrix}$	__[−cns]	__[+sib]	$\_\begin{bmatrix}+\text{cns}\\-\text{sib}\end{bmatrix}$	__[−cns]
[+sib]__	$\frac{0}{11}$ (0.0)	——	$\frac{4}{7}$ (0.57)	$\frac{0}{7}$ (0.0)	$\frac{1}{2}$ (0.5)	$\frac{6}{6}$ (1.0)
$\begin{bmatrix}+\text{cns}\\-\text{sib}\end{bmatrix}\_$	$\frac{1}{12}$ (0.08)	$\frac{2}{2}$ (1.0)	$\frac{8}{8}$ (1.0)	$\frac{3}{15}$ (0.2)	$\frac{0}{2}$ (0.0)	$\frac{11}{11}$ (1.0)
[−cns]__	$\frac{6}{15}$ (0.4)	$\frac{4}{7}$ (0.57)	$\frac{6}{8}$ (0.75)	$\frac{20}{31}$ (0.64)	$\frac{17}{20}$ (0.85)	$\frac{37}{37}$ (1.0)
	PROFESSIONAL-CLASS MEN			PROFESSIONAL-CLASS WOMEN		
[+sib]__	$\frac{3}{4}$ (0.75)	$\frac{7}{7}$ (1.0)	$\frac{9}{10}$ (0.9)	$\frac{11}{24}$ (0.47)	$\frac{4}{4}$ (1.0)	$\frac{14}{15}$ (0.93)
$\begin{bmatrix}+\text{cns}\\-\text{sib}\end{bmatrix}\_$	$\frac{5}{7}$ (0.71)	$\frac{5}{6}$ (0.83)	$\frac{9}{9}$ (1.0)	$\frac{5}{5}$ (0.56)	$\frac{5}{6}$ (0.83)	$\frac{9}{9}$ (1.0)
[−cns]__	$\frac{26}{26}$ (1.0)	$\frac{11}{12}$ (0.92)	$\frac{16}{16}$ (1.0)	$\frac{20}{21}$ (0.95)	$\frac{15}{16}$ (0.93)	$\frac{42}{42}$ (1.0)

TABLE 5. Presence of *que* in complement over total number of complements in nine phonological environments, for four speakers in each category. Relative frequencies in parentheses.

The presence or absence of *que* is highly dependent on the nature of the preceding and following phonological segments; and the three-way distinction among sibilants,[9] other consonants, and vowels is most relevant in this respect. In addition, there are strong effects of the occupational class of the informants. Analysis of the first half-hour of each informant's recorded interview produced the distribution of *que* summarized in Table 5. The data are cross-classified by preceding and following phonological environments, occupational class, and sex of informant.

The data in Table 5 were used to estimate the effect of each feature on the operation of a *que* deletion rule in MF, postulated to be of this form:

$$\text{(24)}\quad que \longrightarrow \langle\emptyset\rangle \;/ \left(\begin{matrix}[+\text{sib}]\\ \begin{bmatrix}+\text{cns}\\-\text{sib}\end{bmatrix}\\ [-\text{cns}]\end{matrix}\right) \#\# \begin{bmatrix}\overline{\quad\quad}\\+\text{compl}\end{bmatrix} \#\# \left(\begin{matrix}[+\text{sib}]\\ \begin{bmatrix}+\text{cns}\\-\text{sib}\end{bmatrix}\\ [-\text{cns}]\end{matrix}\right)$$

In this instance a multiplicative application probabilities model was used; this was far more consistent with the data than a non-application model, as measured by a chi-square comparison of predicted vs. observed frequencies. This results in a clear picture of the interplay of linguistic and social factors, as summarized in Table 6. Preceding and following sibilants freely permit *que* deletion; but the absence of [+sib] and [+cns] tend progressively to restrict rule application, so that preceding and following vowels make for very conservative environments. This phenomenon strongly resembles Labov's rule (1969:748, rule 14) for the simplification of *-sK* clusters. The occupational characteristics of a speaker also tend to affect the rule;

[9] We use the abbreviation [+sib] = [+cor, +cont, −son].

Input probability $p_0 = 1$.

Preceding environment:	[+sib]__	$\begin{bmatrix}+\text{cns}\\-\text{sib}\end{bmatrix}$__	[−cns]__
Effect:	1	0.85	0.37
Following environment:	__[+sib]	__$\begin{bmatrix}+\text{cns}\\-\text{sib}\end{bmatrix}$	__[−cns]
Effect:	1	0.50	0.10
Occupational class:	workers	professional	
Effect:	1	0.35	
Sex:	women	men	
Effect:	1	1	

TABLE 6. Effect of each factor affecting *que* deletion in MF, according to a multiplicative application probabilities model.

working-class informants tend to delete the variable more frequently. On the other hand, the sex of the informant appears to have no effect on the operation of the rule.

Table 7 shows the fit between observed frequencies of *que* deletion and those predicted by the model. Because of the small number of cases (characteristic of syntactic, in contrast to phonological, data) it is not possible to do a chi-square check on how well this model predicts individual performances, as we did with the Panamanian *r*-rule. Bickerton 1973 has re-analysed these *que* data, as presented by G. Sankoff (MS), in terms of an implicational scale of individual grammars. Again, because of the low number of sentences per individual, it is not clear how many, if any, individuals can be convincingly fitted into his scale. The statistical problem of fairly and rigorously distinguishing between scaling and variable rule analyses on the basis of small data sets is discussed by D. Sankoff & Rousseau (MS).

Note that in our analysis [+sib] is a subcategory of [+cns], hence [−cns] is a subcategory of [−sib]. The mutual exclusivity condition for the features in a list

	WORKING-CLASS MEN			WORKING-CLASS WOMEN		
	__[+sib]	__$\begin{bmatrix}+\text{cns}\\-\text{sib}\end{bmatrix}$	__[−cns]	__[+sib]	__$\begin{bmatrix}+\text{cns}\\-\text{sib}\end{bmatrix}$	__[−cns]
[+sib]__	0	—	6.3	0	1.0	5.4
	(0)		(4)	(0)	(1)	(6)
$\begin{bmatrix}+\text{cns}\\-\text{sib}\end{bmatrix}$__	1.8	1.2	7.3	2.2	1.2	10.0
	(1)	(2)	(8)	(3)	(0)	(11)
[−cns]__	9.5	5.7	7.7	19.6	16.4	35.6
	(6)	(4)	(6)	(20)	(17)	(37)
	PROFESSIONAL-CLASS MEN			PROFESSIONAL-CLASS WOMEN		
[+sib]__	2.6	5.8	9.6	15.5	3.3	14.5
	(3)	(7)	(9)	(11)	(4)	(14)
$\begin{bmatrix}+\text{cns}\\-\text{sib}\end{bmatrix}$__	4.9	5.1	8.7	6.3	5.1	8.7
	(5)	(5)	(9)	(5)	(5)	(9)
[−cns]__	22.6	11.2	15.8	18.3	15.0	41.4
	(26)	(11)	(16)	(20)	(15)	(42)

TABLE 7. Frequency of presence of *que* in complements predicted by multiplicative application probabilities model for rule 24, compared with observed frequencies. See Table 5 for totals.

may be relaxed as described in §6 above, so that the lists for preceding and following environments could both be written ⟨[+sib], [+cns]⟩ as in 8b; thus the effects of [+sib], [−sib], [+cns], and [−cns] respectively become 1, 0.85, 1, 0.44 in preceding position, and 1, 0.5, 1, 0.2 in following position.

11. Contraction and deletion of the copula in Black English (NNE). As part of his 1969 analysis of the reduction of the copula in NNE, Labov postulates that, as a first step, it undergoes contraction according to this variable rule:

$$(25)\quad \begin{bmatrix} +\text{voc} \\ -\text{str} \\ +\text{cen} \end{bmatrix} \rightarrow (\emptyset) \,/\, \begin{bmatrix} {}^{*}\text{pro} \\ \alpha\text{V} \end{bmatrix} \#\# \begin{bmatrix} \rule{1em}{0.4pt} \\ +\text{T} \end{bmatrix} \underset{[{}^{*}\text{nas}]}{C_0^1} \#\# \begin{bmatrix} \alpha\text{Vb} \\ \beta\text{gn} \\ -\gamma\text{NP} \end{bmatrix}$$

It is then deleted according to this rule:

$$(26)\quad [+\text{cons}] \rightarrow (\emptyset) \,/\, \begin{bmatrix} -\beta\text{V} \\ \gamma\text{pro} \\ {}^{*}\text{strid} \end{bmatrix} \#\# \begin{bmatrix} \rule{1em}{0.4pt} \\ -\text{nas} \\ +\text{cont} \end{bmatrix} \#\# \begin{bmatrix} \alpha\text{Vb} \\ \beta\text{gn} \\ -\gamma\text{NP} \end{bmatrix}$$

The information incorporated in these rules was extracted from Labov's Table 6, summarized in our Tables 8 and 9.

	__NP	__PA-Loc	__Vb	__gn
Pro__				
[−cns]	$\frac{2}{32}$ (0.06)	$\frac{1}{65}$ (0.02)	$\frac{1}{34}$ (0.03)	$\frac{0}{23}$ (0.0)
Other NP__				
[+cns]	$\frac{22}{35}$ (0.63)	$\frac{24}{32}$ (0.75)	$\frac{5}{14}$ (0.36)	$\frac{1}{9}$ (0.11)
[−cns]	$\frac{13}{64}$ (0.20)	$\frac{7}{23}$ (0.30)	$\frac{2}{14}$ (0.14)	$\frac{0}{6}$ (0.0)

Table 8. (From Labov's Table 6, corrected.) Frequency of non-contracted copula over total number of cases, by preceding and following environment. Relative frequencies in parentheses.

	__NP	__PA-Loc	__Vb	__gn
Pro__				
[−cns]	$\frac{18}{30}$ (0.6)	$\frac{28}{64}$ (0.44)	$\frac{7}{33}$ (0.21)	$\frac{1}{23}$ (0.04)
Other NP__				
[+cns]	$\frac{5}{13}$ (0.38)	$\frac{4}{8}$ (0.50)	$\frac{0}{9}$ (0.0)	$\frac{1}{8}$ (0.13)
[−cns]	$\frac{36}{51}$ (0.72)	$\frac{10}{16}$ (0.63)	$\frac{8}{12}$ (0.67)	$\frac{0}{6}$ (0.0)

Table 9. Frequency of non-deleted copula over total number of contracted copula. Relative frequencies in parentheses.

We used these data in estimating the parameters of multiplicative non-application probabilities models for contraction and deletion. These predicted frequencies are very similar to the observations. Details are presented in Tables 10–13.

Our systematic use of variable constraints enables us to compare feature effects across different rules and, in general, to achieve Labov's original aim in introducing variable rules: the demonstration of important quantitative linguistic relations.

Input probability $p_0 = 0.25$.

Preceding NP:	Pro__	Other NP__		
Effect:	0.86	0		
Preceding segment:	[+cns]__	[−cns]__		
Effect:	0	0.65		
Following environment:	__NP	__PA-Loc	__Vb	__gn
Effect:	0.16	0	0.49	0.89

TABLE 10. Effects of variable constraints on contraction rule.

Input probability: $p_0 = 0.27$.

Preceding NP:	Pro__	Other NP__		
Effect:	0.29	0		
Preceding segment:	[+cns]__	[−cns]__		
Effect:	0.42	0		
Following environment:	__NP	__PA-Loc	__Vb	__gn
Effect:	0	0.13	0.44	0.90

TABLE 11. Effects of variable constraints on deletion rule.

	__NP	__PA-Loc	__Vb	__gn
Pro__	1.0	2.3	0.6	0.1
[−cns]	(2)	(1)	(1)	(0)
Other NP__				
[+cns]	21.9	24.0	5.3	0.8
	(22)	(24)	(5)	(1)
[−cns]	14.0	6.0	1.9	0.2
	(13)	(7)	(2)	(0)

TABLE 12. Observed frequencies of contraction (in parentheses), compared to predictions of multiplicative model. See Table 8 for totals.

	__NP	__PA-Loc	__Vb	__gn
Pro__	15.7	28.9	9.2	1.2
[−cns]	(18)	(28)	(7)	(1)
Other NP__				
[+cns]	5.1	2.7	1.9	0.3
	(5)	(4)	(0)	(1)
[−cns]	37.3	10.1	4.7	0.4
	(36)	(10)	(8)	(0)

TABLE 13. Observed frequencies of deletion (in parentheses), compared to predictions of multiplicative model. See Table 9 for totals.

His data indicate that contraction and deletion are similarly affected by the grammatical environment, but predictably opposed in the phonological one. These effects are confirmed concisely and clearly in our Tables 10–11, where a following verb has an almost identical effect in favoring contraction and deletion, while the roles of preceding vowels and consonants are reversed in the two rules.[10] The effect of Pro__ in favoring application of both rules, and the parallel lack of any clear distinction between __NP and __PA-Loc further confirms this relation.

Rewriting rules 25–26 in our notation, we retain Labov's asterisk notation for obligatory environments:

$$(27)\ \begin{bmatrix} +\text{voc} \\ -\text{str} \\ +\text{cen} \end{bmatrix} \longrightarrow \langle\emptyset\rangle \,/\, \begin{matrix} \langle\text{Pro}\rangle \\ \langle[-\text{cns}]\rangle \end{matrix}\ \#\#\ \begin{bmatrix} \_\_\_ \\ +\text{T} \end{bmatrix} \underset{[*+\text{nas}]}{C_0^1}\ \#\#\ \left(\begin{matrix} \text{Vb} \\ \text{gn} \\ \text{NP} \\ \text{PA-Loc} \end{matrix} \right)$$

$$(28)\ [+\text{cns}] \longrightarrow \langle\emptyset\rangle \,/\, \begin{matrix} \langle\text{Pro}\rangle \\ \langle+\text{cns}\rangle \\ \langle*+\text{strid}\rangle \end{matrix}\ \#\#\ \begin{bmatrix} \_\_\_ \\ [-\text{nas}] \\ [+\text{cont}] \end{bmatrix} \#\#\ \left(\begin{matrix} \text{Vb} \\ \text{gn} \\ \text{NP} \\ \text{PA-Loc} \end{matrix} \right)$$

Then in rule 27, e.g., the environment / ## [___, +T] [C, +nas] ## is one in which the rule always applies. Note that in 27 [+T] denotes a feature which is REQUIRED for the rule to be applicable at all, and [* + nas] denotes a feature which OBLIGES the rule to apply (as long as all other required features are present, of course).[11]

12. DISCUSSION AND SUMMARY. The full importance of variable rules can be appreciated only from a certain paradigmatic standpoint, one which constitutes a slight but distinct shift from most generative theory. First, this view requires that the notion of competence must be strengthened to include representation of systematic covariation between elements of language, even when this covariation cannot be described in categorical (zero–one) terms. The representation should be in the form of abstract probabilities, which measure tendencies for rule application in various environments. Second, this viewpoint focuses on variation as it exists in unedited performance, minimally influenced by the observer-linguist, and recorded from the behavior of some well-defined sample of speakers in specific speech contexts. In addition to these two basic characteristics, studies within this paradigm usually involve languages in which much is already known about the grammar and dialectal variation, and speech communities for which there is interest in variation in language use related to style, socio-economic class, age, etc. (G. Sankoff, MS).

[10] A preceding vowel favors the loss of the vowel from the copula; but given that contraction has occurred, it is a preceding consonant that favors the deletion of the remaining consonant.

[11] In the multiplicative non-application probabilities model, the obligatory effect of a feature i can also be indicated by $p_i = 1$; and in the application probabilities model, the minimally required nature of a binary feature j can also be indicated by $p_{\sim j} = 0$. This can be verified by substitution in formulae 2 and 4 respectively.

The power of this approach lies in the uniquely well-defined and economical relationship which it posits between competence and linguistic performance, analogous to that between a probability distribution and a sample, or between a model and a simulation.[12] This relationship not only integrates generative and behavioral aspects in an elegant way, but is also easily operationalized so as to provide consistent and reproducible results. The utility of a theory containing variable rules is magnified many times as a consequence of the ease and naturalness with which it extends from purely linguistic applicability to the domains of sociolinguistics, stylistics, dialectology, and language change, individually or in combination.

In this paper, we have tried to clarify some of the probabilistic aspects left unresolved in Labov's work. We distinguish rule probabilities from rule frequencies, assigning the former to competence and the latter to performance. We have reformulated the theory of interaction between environmental features in influencing rule probabilities, and we hypothesize that the universally 'unmarked' mode of interaction in variable rules is that of independence, as opposed to synergism or antagonism. This hypothesis seems preferable to Labov's original additive model for variable constraints, on the basis of several criteria.

On the practical side, we have developed a systematic methodology for treating observed frequency data associated with variable rules, and estimating the appropriate probabilistic parameters. This produces meaningful results for data on *r*-spirantization in Panamanian Spanish, complementizer *que* deletion in Montreal French, and contraction and deletion of the copula in Black English. Features of this methodology include the ability to detect instances of non-independent interaction between environmental factors, and flexibility in incorporating successive refinements in linguistic and non-linguistic conditioning (Cedergren 1973a).

By incorporating a speaker's non-linguistic parameters (class, sex, age etc.) in the input probability, and by setting the strictly linguistic parameters at fixed values for the whole community, this approach neatly solves the problem of community heterogeneity—perhaps too neatly; care should be taken to detect categorical rule differences where these exist (cf. Bickerton 1971). We have written our rules in terms of individual speakers' grammars, and not in terms of group or social grammars; but these constructs can be attained if desired by appropriate adjustments of input probabilities. In the one case in which there are enough data to judge, i.e. for Panamanian *r*-spirantization, the procedure of calculating individual parameters from aggregate data proves warranted. Of course, when complete individual data are available, these can be put directly into our parameter estimation procedure. Further statistical methods must be developed in order to judge when small data sets on individual speakers can be aggregated without obscuring categorical distinctions between individual grammars.

One of the most positive features of variable rule analysis is the convergence in the results of research independently carried out in related speech communities. This can be seen in comparing Labov's work in New York with the Detroit study by Shuy, Riley, and Wolfram (cf. Wolfram 1969) and other studies of Black Eng-

[12] For related work at the level of phrase-structure grammars, see Suppes 1970, D. Sankoff 1971a; at the lexical level, Lehrer 1970, D. Sankoff 1971b, Labov 1973c.

lish, or in the comparison between Panamanian Spanish (Cedergren 1973a) and Puerto Rican in New York (Ma & Herasimchuk 1968). The models suggested here appear to fit in with the aims of the original research in this field and to provide a jointly satisfactory framework (Labov 1973a, b) for further work, in facilitating the comparison of variable rules and contributing to the establishment of universal laws for variable constraints.[13]

REFERENCES

BAILEY, CHARLES-JAMES N. 1971. Trying to talk in the new paradigm. Papers in Linguistics 4.312–38.

——, & R. W. SHUY (eds.) 1973. New ways of analyzing variation in English. Washington, D.C.: Georgetown University Press.

BICKERTON, DEREK. 1971. Inherent variability and variable rules. Foundations of Language 7.457–92.

——. 1973. Quantitative versus dynamic paradigms: the case of Montreal *que*. In Bailey & Shuy, 23–43.

CANFIELD, D. LINCOLN. 1962. La pronunciación del español en América. Bogotá: I.C.C.

CEDERGREN, HENRIETTA. 1973a. On the nature of variable constraints. In Bailey & Shuy, 13–22.

——. 1973b. Interplay of social and linguistic factors in Panama. Cornell dissertation.

DECAMP, DAVID. 1971. Toward a generative analysis of a post-creole speech continuum. Pidginization and creolization of languages, ed. by D. Hymes, 349–70. Cambridge: University Press.

ELLIOT, DALE; STANLEY LEGUM; & SANDRA A. THOMPSON. 1969. Syntactic variation as linguistic data. Papers from the 5th Regional Meeting, Chicago Linguistic Society, 52–9.

ESTES, W. K. 1964. Probability learning. Categories of human learning, ed. by A. W. Melton, 90–128. New York: Academic Press.

KLEIN, SHELDON. 1965. Control of style with a generative grammar. Lg. 41.619–31.

KUHN, THOMAS S. 1970. The structure of scientific revolutions. Chicago: University of Chicago Press.

LABOV, WILLIAM. 1969. Contraction, deletion and inherent variability of the English copula. Lg. 45.715–62.

——. 1971. Methodology. A survey of linguistic science, ed. by William O. Dingwall, 412–97. College Park: University of Maryland Linguistics Program.

——. 1973a. Sociolinguistic patterns. Philadelphia: University of Pennsylvania Press.

——. 1973b. Language in the Inner City. Philadelphia: University of Pennsylvania Press.

——. 1973c. The meanings of words and their boundaries. In Bailey & Shuy, 340–73.

LEE, W. 1971. Decision theory and human behavior. New York: Wiley.

LEHRER, ADRIENNE. 1970. Indeterminacy in semantic description. Glossa 4.87–110.

MA, ROXANA, & ELEANOR HERASIMCHUK. 1968. The linguistic dimensions of a bilingual neighborhood. Bilingualism in the Barrio, ed. by J. Fishman et al., 349–464. Bloomington: Indiana University.

NAVARRO-TOMÁS, TOMÁS. 1963. Manual de pronunciación española. New York: Hafner.

RICORD, E. 1971. El español de Panamá. Panamá: University of Panama.

SANKOFF, DAVID. 1971a. Branching processes with terminal types: application to context-free grammars. Journal of Applied Probability 8.233–40.

——. 1971b. Dictionary structure and probability measures. Information and Control 19.104–13.

[13] We would like to thank J. B. Kruskal for a detailed critique of an earlier version of this paper, and William Labov for his collaboration and guidance through the various revisions.

——, & PASCALE ROUSSEAU. MS. A method for assessing variable rule and implicational scale analyses of linguistic variation. To appear, Proceedings of the International Conference on Computers in the Humanities, ed. by L. Mitchell, Edinburgh University Press.

——, & GILLIAN SANKOFF. 1973. Sample survey methods and computer assisted analysis in the study of grammatical variation. Canadian languages in their social context, ed. by R. Darnell, 7–63. Edmonton: Linguistic Research, Inc.

SANKOFF, GILLIAN. MS. A quantitative paradigm for studying communicative competence. Conference on the Ethnography of Speaking, Austin, Texas.

——, & HENRIETTA J. CEDERGREN. 1972. Sociolinguistic research on French in Montreal. Language in Society 1.173–4.

SUPPES, PATRICK. 1970. Probabilistic grammars for natural languages. Synthèse 22.95–116.

TUCKER, G. R. 1967. French speakers' skill with grammatical gender. McGill University dissertation.

WEINREICH, URIEL; WILLIAM LABOV; & MARVIN I. HERZOG. 1968. Empirical foundations for a theory of language change. Directions for historical linguistics, ed. by W. P. Lehmann & Y. Malkiel, 97–195. Austin: University of Texas Press.

WOLFRAM, WALTER A. 1969. A sociolinguistic description of Detroit Negro speech. Washington, D.C.: Center for Applied Linguistics.

ZAMORA VICENTE, A. 1967. Dialectología española. Madrid: Editorial Gredos.

[Received 7 March 1973]