Language Variation and Change, 1 (1989), 1-18. Printed in the U.S.A.

Statistical evidence for rule ordering

DAVID SANKOFF
Centre de recherches mathématiques
Université de Montréal

PASCALE ROUSSEAU
Département de mathématiques et informatique
Université du Québec à Montréal

ABSTRACT

A set of ordered rules for generating variants of a variable determines (a) underlying/surface distinctions among some of the variants and (b) a hierarchical classification of the variants. In the analytical framework of variable rules, frequency data on variant occurrences in context bear only on (b) and not on (a). We provide a combinatorial characterization and enumeration of the set of classifications on *n* variants, the set of underlying/surface configurations, and the set of rule orders. We describe the statistical and computational techniques for generalizing variable rule analysis to the inference of rule order. These procedures are applied to the problems of the reduction of syllable-final consonants ⟨s⟩, ⟨n⟩, and ⟨r⟩ in Caribbean Spanish ($n = 3$, 4, 6 variants, respectively). Previous analyses have tended to assume that successive weakenings occur in an intrinsic order determined by phonological strength. Our results show that aspiration and deletion can indeed be seen to be intrinsically ordered in both ⟨s⟩ and ⟨r⟩ reduction, though an unordered analysis is equally likely in the case of ⟨s⟩. On the other hand, velarization and deletion of ⟨n⟩ are unordered, while vocalization is a subsequent process, independent of the other two. Similarly, spirantization, aspiration, and lateralization of ⟨r⟩ are unordered, as confirmed by data sets from both Puerto Rican and Panamanian speakers. Furthermore, with both ⟨n⟩ and ⟨r⟩, intrinsically ordered rule schemata proved to be extremely unlikely by statistical criteria. Syllable-final consonant reduction then consists of largely independent processes, most of which occur simultaneously.

INTRODUCTION

When a phonological variable is manifested as one of three or more variants, generating these variants through a series of phonological rules brings up the difficult problem of *rule ordering*. This is closely related, but not identical, to the problem of *distinguishing underlying forms* from intermediate forms and from purely surface variants. In addition to these concerns, partly substantive and partly notational, decisions about rule ordering are also connected to the task of *classifying tokens* for purposes of the statistical analysis of performance data.

These three problem areas - underlying forms, rule orders, and token classification - have often been confused. In this article, we clarify the logical relationship among them and suggest an approach to solving them, as far as possible, based on the statistical principle of maximum likelihood, applied to performance data. It will emerge that these data have no bearing on the underlying/surface distinction, but they bear to a surprising degree on the rule order and token classification problems; more specifically, the rule order problem can be decomposed into two problems: that of underlying/surface distinctions, for which the data have no relevance, and that of token classification, whose solution the data determine completely, in a statistical sense.

VARIABLE RULES

The rule in **(1)** gives a rough account of ⟨l⟩ deletion in subject pronouns in Montreal French: that is, pronouns like *il* 'he/it', *elle* 'she', and *ils* 'they' (for details, see Poplack & Walker, 1986; Pupier & Légaré, 1973; G. Sankoff & Cedergren, 1971; Santerre, Noiseux, & Ostiguy, 1977).

$$l \rightarrow \langle \emptyset \rangle \;/\; \#\; V \begin{bmatrix} \underline{\qquad} \\ \langle +\text{sing} \rangle \\ \langle +\text{masc} \rangle \end{bmatrix} \# \; \langle +\text{cons} \rangle \tag{1}$$

The rule says that ⟨l⟩ is deleted variably and that the features in the environment that favour its deletion are: a following consonant, in contrast to a following vowel; the ⟨l⟩ being in a singular pronoun, *il va* 'he goes' rather than *ils vont* 'they go'; and the pronoun having masculine gender (*il* vs. *elle*).

Now, the basic working hypothesis in the statistical study of performance is that there is some set of numbers associated with the set of features pertinent to the rule, that is, there is a number associated with the feature *consonant,* and another with *vowel,* a number each with *singular, masculine,* and so on, and that for any particular token we can derive a probability that the ⟨l⟩ is going to be deleted based on the numbers associated with those features actually present in the context of the token. Consider, for example, the ⟨l⟩ in *elle va* 'she goes'. By our postulate, there is some number f associated with feminine (−masc) pronouns, another number s associated with singular pronouns, and yet another v, with a following vowel (−cons) environment, and in every conjuncture of these three conditions, the three numbers are combined in the way summarized in **(2)** to give the probability p that the ⟨l⟩ will be deleted:

$$\log[p/(1-p)] = f + s + v \tag{2}$$

We use $\log p/(1-p)$ on the left-hand side of Formula (2), instead of just p, to ensure that no matter how large or small, negative or positive, are the values of parameters like f, s, v, the probability p will never be less than 0 or more than 1.

In other contexts, other parameters will be necessary instead of one or more of those shown in (2), such as m, t, or c for masculinity, plurality (−sing), or following consonant. It should be stressed that each parameter has a fixed value in all contexts where it is pertinent; it is just the different combinations of parameters that result in different deletion probabilities.

Different analyses may postulate differing sets of relevant *factors*, or *constraints*, on the linguistic variable. And aside from linguistic factors, such as number, gender, and phonological environment, there may be one or more parameters representing the overall deletion tendencies of the speaker or the social grouping(s) into which the speaker may be classified or the speech style or context of the utterance.

This basic hypothesis is known as the *additivity of constraints* hypothesis because of the addition of the parameters in (2). Some set of constraints can usually be found that account for the data in an additive way, though sometimes this involves the introduction of additional parameters into the formula to account for statistical interaction between the factors.

Before quantitative analysis of performance data, we cannot know the numerical value of the parameters, and hence we cannot evaluate the relative importance of the different factors on ⟨l⟩ deletion. The first step, then, is to collect a large number of tokens in some corpus of natural speech. We separate the tokens occurring in our corpus into two classes—those with an audible articulation of the ⟨l⟩, and those without—for each context in which the pronoun appeared. We then code each context according to the presence or absence of each of the relevant factors in its context. All this information can then be fed into a statistical procedure in order to estimate the parameters. The particular procedure most often used in variable rule analysis (D. Sankoff, 1988) is known as a logistic regression or the logit model.

In carrying out the estimation of the parameters, or constraint effects, the procedure is guided by the criterion of *maximum likelihood*. This ensures that the set of parameter estimates obtained is that for which the observed deletion data are most likely to have been generated, in the statistical sense, if the hypothesis of additivity is correct. The *likelihood* of an analysis is a numerical measure of the fit between the data and the set of parameter values to be used in Formula (2).

RULE DIRECTIONALITY

It is important to note that the data, tokens divided into two groups—[l]s and ∅s in each context—do not bear on the directionality of the rule, that is, on the underlying/surface distinction. We could just as well analyze the data as being generated by the reinsertion rule:

$$\varnothing \rightarrow \langle l \rangle \; / \; \# \; V \begin{bmatrix} \_\_\_\_ \\ \langle -\text{masc} \rangle \\ \langle -\text{sing} \rangle \end{bmatrix} \# \; \langle -\text{cons} \rangle \tag{3}$$

which some would claim to be a more accurate characterization of Montreal French. Here, the statistical analysis of the data will always estimate each binary feature effect to be equal to the effect of its opposite were the data to be analyzed as in Rule **(1)**. In addition, the likelihoods of the two analyses will be identical. Thus, statistically speaking, a rule and its opposite are always equally plausible for a given data set. Note that these two rules share the same classification of tokens into two groups: [l]s and ∅s. As should become increasingly clear, the likelihood is associated with the token classification, not the direction of the rule. This irrelevance of the data for rule directionality or for the underlying/surface distinction is a trivial example of a fact of much greater consequence in the case where a linguistic variable can be realized by three or more variants instead of just two.

THE THREE-VARIANT CASE

The syllable-final ⟨s⟩ variable in Caribbean Spanish generally has three manifestations denoted by [s], [h], and ∅. These variants could conceivably be generated by any of the 15 different pairs of ordered or unordered rules in Table 1. Indeed, many of these rule order schemata have been proposed at one time or another.

In Table 1, the 15 pairs of rules are grouped into three types under the column headings: intrinsic order, extrinsic order, and unordered. For each analysis we present, from left to right, (i) the inherent assumptions about which forms underlie which other forms, using the ">" notation; (ii) the rule order schema; and (iii) the token classification, as explained in Table 2.

Thus, in the rule order for the analysis in the upper left of the table,

$$\begin{array}{c} s \rightarrow h \\ h \rightarrow \varnothing \end{array}$$

tokens of ∅ are assumed to result from successive applications of an aspiration rule and a deletion rule, while in the center column

$$\begin{array}{c} s \rightarrow h \\ s \rightarrow \varnothing \end{array}$$

tokens of ∅ are assumed to result from nonapplication of an aspiration rule followed by the application of a deletion rule. The entry in the lower left of the table

$$\begin{array}{c} \varnothing \rightarrow s \\ s \rightarrow h \end{array}$$

TABLE 1. *Relationship between underlying form assumptions, rule order schemata, and token classification for statistical analysis*[a]

Assumed Underlying Form	Intrinsic Order	Token Classification	Assumed Underlying Form	Extrinsic Order	Token Classification	Assumed Underlying Form	Unordered	Token Classification
s > h > ∅	s → h h → ∅	B	s	s → h s → ∅	C	s	{s → h, s → ∅}	A
s > ∅ > h	s → ∅ ∅ → h	B	s	s → ∅ s → h	D			
h > s > ∅	h → s s → ∅	C	h	h → s h → ∅	B	h	{h → s, h → ∅}	A
h > ∅ > s	h → ∅ ∅ → s	C	h	h → ∅ h → s	D			
∅ > h > s	∅ → h h → s	D	∅	∅ → h ∅ → s	C	∅	{∅ → s, ∅ → h}	A
∅ > s > h	∅ → s s → h	D	∅	∅ → s ∅ → h	B			

[a]Unordered pairs of rules in braces are applied simultaneously; otherwise upper rule is applied first, followed by the lower rule. See Table 2 for explanation of token classifications A, B, C, and D.

TABLE 2. *Four ways of generating the distinctions among three variants*[a]

A	B	C	D
{s}, {h}, {∅}	1. {s}, {h, ∅} 2. {h}, {∅}	1. {h}, {s, ∅} 2. {s}, {∅}	1. {∅}, {s, h} 2. {s}, {h}

[a]Classification A effects all distinctions at once. Classifications B, C, and D distinguish one variant from the other two first, then distinguish between these latter two as a second step.

shows [h] tokens as deriving from an [s] insertion rule, followed by an aspiration rule. In the middle of the right-hand column,

$$\begin{Bmatrix} h \rightarrow s \\ h \rightarrow \varnothing \end{Bmatrix}$$

we have ∅ and [s] being generated in an unordered way from an underlying [h], by simultaneous processes of deletion and, say, normative correction.

As in the two-variant case, the data do not necessarily bear on all the distinctions between the various possible linguistic analyses (i.e., rule order schemata). In fact, as shown in Table 2, there are only four different ways of statistically analyzing the data, corresponding to the four possible ways of generating the distinction among the three variants. One way, A, involves dividing the tokens directly into three classes and performing a three-variant variable rule analysis (Kemp, 1979; D. Sankoff & Labov, 1979). The other three ways, B, C, and D, involve first a division of one variant versus the other two combined, followed by the division of the latter two types of tokens. At each of these two steps, an ordinary two-variant variable rule analysis is performed.

Table 1 illustrates completely, for the three-variant case, the relationship between rule order, underlying form assumptions, and token classification. Rule order clearly determines token classification and something about underlying forms, though it does not necessarily determine deep/surface distinctions between each pair of variants. Token classification alone determines neither rule order nor underlying form. Underlying form assumptions alone determine neither rule order nor token classification. But token classification together with underlying distinctions do uniquely determine the rule order. An algorithm for generating all possible rule order schemata will be described in the next section, and this will be applied to the four-variant and six-variant cases in the ensuing sections.

In a study of data from Panama City, ⟨s⟩ reduction was analyzed according to token classifications A, B, C, and D (Cedergren, Rousseau, & Sankoff, 1986). The four different analyses could then be compared by means of their log likelihoods.[1] To find the log likelihood of a multi-rule analysis, such as those represented by B, C, or D, the log likelihoods of the different steps are simply added together. The mathematical properties of maximum likelihood token classifications are investigated in Rousseau and Sankoff (1989).

Classification B in Table 2 proved to be much more likely than C or D. Under the historically, comparatively, and phonologically motivated assumption of underlying [s] and intermediate [h], this vindicated an intrinsically ordered analysis versus the two possible extrinsic orders.[2]

However, the *unordered* analysis in A, which involves two of the forms ([h] and ∅) being simultaneously generated from the third, [s], had a likelihood equal to that of the intrinsic order, so that a choice between the intrinsic and unordered approaches could not seem to be justified based on these data. (Without the assumption of underlying [s] and intermediate [h], of course, we have succeeded only in reducing the choices in Table 1 from 15 to 7.)

THE SET OF POSSIBLE RULE SCHEMATA

If, rather than two or three variants, there are n to be generated, then each rule order schema must have $n - 1$ rules. The first rule may have any variant on the left and any other on the right of the arrow. The second rule may have either of the variants already mentioned on the left and any previously unmentioned variant on the right, and this criterion is applicable for all succeeding rules: any previously mentioned term on the left, any previously unmentioned term on the right.

This construction generates all possible schemata of ordered rules. Each such schema must then be examined to see if it can be modified to produce one or more schemata containing unordered rules: If two or more consecutive rules in a rule order have the same left-hand side, they may be unordered by adding braces, as in the right-hand column of Table 1.

Three remarks should be made about this procedure.

(i) Though it produces all, and only those, possible rule schemata, it does not constitute a way of counting these schemata, as each schema containing unordered rules will be generated in at least two different ways. The Appendix contains exact counting procedures for rules, schemata, classifications, and possible sets of underlying form configurations. Table 3 depicts the results for n up to 8.

(ii) If two or more consecutive rules in a schema do not involve any of the same variants, there will be other schemata generated containing these same consecutive rules but in a different order. These schemata look different but are substantively the same.[3] We may enclose such subsets of rules in braces, or display them in separate columns, to indicate the lack of any necessary

TABLE 3. *Counts of possible analyses for up to eight variants*[a]

	R_n^*	T_n	N_n	M_n
		Number of Different:		
n	Rules Involving at Most *n* Variants (not counting directionality) e.g., {u, x}, {v, y}	Token Classifications e.g., {u, x}, {v, y} {u}, {x} {y}, {v}	Sets of Underlying Assumptions e.g., x > v > y x > u	Rule Order Schemata e.g., x → v {x → u, v → y}
2	1	1	2	2
3	7	4	9	15
4	36	26	64	184
5	171	236	625	3,155
6	813	2,752	7,776	69,516
7	4,012	39,208	117,649	1,871,583
8	20,891	660,032	2,097,152	59,542,064

[a]See Tables 1 and 2 for case $n = 3$, and the Appendix for the general method of calculation.

ordering among them, but we will reserve the term "unordered" for the case where several variants are simultaneously generated from another.

(iii) We do not consider the possibility that a single variant may be generated in more than one way *(convergent rules)* within a given schema, as in (4).

$$\begin{aligned} s &\rightarrow \varnothing \\ s &\rightarrow h \\ h &\rightarrow \varnothing \end{aligned} \qquad (4)$$

This could be a serious shortcoming with certain variables. Statistical methods for detecting and handling such *hidden variants* are being developed (D. Sankoff & Rousseau, 1982).

Given a rule order schema, how can we deduce the underlying form assumptions and the token classification associated with it? The former is much easier than the latter. Any variant on the left of a rule is "deeper" than that on the right, and "depth" is transitive—if x is deeper than y and y is deeper than z, then x is deeper than z. The procedure for generating rule order schemata ensures that no contradictions (x is deeper than x) may arise. As for token classifications, these may be constructed by first looking at the last rule, then the second-to-last, and so on. If the last rule is $x \rightarrow y$, then the last step in the classification is {x}, {y}, and in all earlier steps, whenever x appears in a group, y must also be there. So that if the second-to-last rule is $z \rightarrow x$ or $x \rightarrow z$, then the second-to-last step in the classification is {z}, {x, y}. Then in all earlier steps, whenever one of x, y, or z appears, they all must appear in the same group. This principle suffices to reconstruct the token classification, keeping in mind that when we encounter a series of unordered rules, the corresponding step in the token classification distinguishes among the right-hand variants of each of these rules and between these and their common left-hand variant.

THE FOUR-VARIANT CASE

We now consider word-final plural verb morpheme ⟨n⟩ in Puerto Rican Spanish, using the data described by Poplack (1979, 1980). Tokens are coded by speaker, following phonological segment, presence of verb irregularity, as well as presence and position of disambiguating plural information elsewhere in the sentence. We denote by N the alveolar variant or cases where the ⟨n⟩ is assimilated to the following consonant; [ŋ], the velarized variant in a nonvelar environment; [Ṽ], the retention of nasality only on the preceding vowel; and ∅, total deletion.

The four variants could conceivably have been generated by any of 184 rule schemata, compared to the 15 possible with three variants and the 2 possible with two variants. In Table 4, we show only the 26 different ways of classifying the tokens, and in Table 5 we present the 26 associated rule order schemata under the assumption that N is the underlying form, [ŋ] is deeper than [Ṽ], and [Ṽ] is deeper than ∅. (Not all these assumptions are necessary for all 26 cases.)

In interpreting results such as those in Table 5, we cannot simply choose the most likely analysis and be confident that this is the "correct" answer. Especially if the log likelihoods of the best few analyses are not very different, we should examine all of these to see if they have points in common. Doing this for Table 5, we see that contrary to traditional successive weakening theories that might correspond to the intrinsic order B, in which the rules apply in the order of a natural weakening process or in stages that can be seen in historical processes, we infer instead that the ⟨n⟩ is either velarized or deleted, by independent processes happening simultaneously or in some order, whereas vocalization never occurs before the other two processes. This is fairly consistent with all the likely analyses and more so with the most likely. It is also confirmed by examining the least likely ones: X, L, R, and so forth, which all tend to derive the nasalized vowel early and tend to derive it before the [ŋ] and before the ∅, in various orders and combinations.

Note that not only are the orders that derive [ŋ] and ∅ before [Ṽ] more likely than those that derive [Ṽ] first or second but those that derive ∅ from [Ṽ] tend to be among the least likely of all. This pattern runs counter to the intrinsic order hypothesis (exemplified by B) made by Cedergren (1973) and Poplack (1979), but are more in accord with the analysis of Poplack (1980), whereby the distinction between ∅ and the other variants is primary.

Another result, not displayed in Table 5, is that the constraint effects in the different analyses tend to be fairly stable. That is, the effect of the following phonological environment and so forth on the various rules tends to

TABLE 4. *Twenty-six token classifications possible with four variants*[a]

	A			
1.	{N}, {ŋ}, {Ṽ}, {∅}			
	B	C	D	E
1.	{N}, {ŋ, Ṽ, ∅}	{N}, {ŋ, Ṽ, ∅}	{N}, {ŋ, Ṽ, ∅}	{N}, {ŋ, Ṽ, ∅}
2.	{ŋ}, {Ṽ, ∅}	{Ṽ}, {ŋ, ∅}	{∅}, {ŋ, Ṽ}	{ŋ}, {Ṽ}, {∅}
3.	{Ṽ}, {∅}	{ŋ}, {∅}	{ŋ}, {Ṽ}	
	F	G	H	I
1.	{ŋ}, {N, Ṽ, ∅}	{ŋ}, {N, Ṽ, ∅}	{ŋ}, {N, Ṽ, ∅}	{ŋ}, {N, Ṽ, ∅}
2.	{N}, {Ṽ, ∅}	{Ṽ}, {N, ∅}	{N}, {Ṽ}, {∅}	{∅}, {N, Ṽ}
3.	{Ṽ}, {∅}	{N}, {∅}		{N}, {Ṽ}
	J	K	L	M
1.	{Ṽ}, {N, ŋ, ∅}	{Ṽ}, {N, ŋ, ∅}	{Ṽ}, {N, ŋ, ∅}	{Ṽ}, {N, ŋ, ∅}
2.	{N}, {ŋ, ∅}	{N}, {ŋ}, {∅}	{ŋ}, {N, ∅}	{∅}, {N, ŋ}
3.	{ŋ}, {∅}		{N}, {∅}	{N}, {ŋ}
	N	O	P	Q
1.	{∅}, {N, ŋ, Ṽ}	{∅}, {N, ŋ, Ṽ}	{∅}, {N, ŋ, Ṽ}	{∅}, {N, ŋ, Ṽ}
2.	{N}, {ŋ}, {Ṽ}	{N}, {ŋ, Ṽ}	{ŋ}, {N, Ṽ}	{Ṽ}, {N, ŋ}
3.		{ŋ}, {Ṽ}	{N}, {Ṽ}	{N}, {ŋ}

	R	S	T
1.	{N, ŋ}, {Ṽ}, {∅}	{N}, {ŋ, Ṽ}, {∅}	{N}, {ŋ}, {Ṽ, ∅}
2.	{N}, {ŋ}	{ŋ}, {Ṽ}	{Ṽ}, {∅}
	U	V	W
1.	{N, Ṽ}, {ŋ}, {∅}	{N, ∅}, {ŋ}, {Ṽ}	{ŋ, ∅}, {N}, {Ṽ}
2.	{N}, {Ṽ}	{N}, {∅}	{ŋ}, {∅}
	X	Y	Z
1.	{N, ŋ}, {Ṽ, ∅}	{N, Ṽ}, {ŋ, ∅}	{N, ∅}, {ŋ, Ṽ}
2a.	{N}, {ŋ}	{N}, {Ṽ}	{N}, {∅}
2b.	{Ṽ}, {∅}	{ŋ}, {∅}	{ŋ}, {Ṽ}

[a]The order of 2a and 2b may be reversed without affecting the results.

be relatively stable from one analysis to another. It does tend to be clearer in the most likely ones and somewhat confused in the less likely ones.

THE SIX-VARIANT CASE

Poplack's (1979) data on final ⟨r⟩ in infinitives in Puerto Rican Spanish involve seven variants, with some of the ⟨r⟩ being realized as [n], especially before pauses or vowels, but the computational effort necessary to handle seven variants is rather excessive at present. For the six-variant case, there are already 2,752 token classifications to consider, each made up of some subset of the 813 possible rules (Table 3), and this requires considerable computer time. For seven variants there are about 39,000 classifications and 4,000 rules, which would not be feasible with the current program. Interest-

TABLE 5. *Rule order schemata under the assumption of increasing depth from ∅ to* [Ṽ] to [ŋ] *to* N[a]

U:	{N → ∅ N → ŋ} N → Ṽ	P:	N → ∅ N → ŋ N → Ṽ	A:	{N → ŋ N → Ṽ N → ∅}	S:	{N → ŋ N → ∅} ŋ → Ṽ	H:	N → ŋ {N → Ṽ N → ∅}
I:	N → ŋ N → ∅ N → Ṽ	D:	N → ŋ ŋ → ∅ ŋ → Ṽ	E:	N → ŋ {ŋ → Ṽ ŋ → ∅}	Z:	N → ŋ {N → ∅ ŋ → Ṽ}	N:	N → ∅ {N → ŋ N → Ṽ}
O:	N → ∅ N → ŋ ŋ → Ṽ	G:	N → ŋ N → Ṽ N → ∅	K:	N → Ṽ {N → ŋ N → ∅}	F:	N → ŋ N → Ṽ Ṽ → ∅	Y:	N → ŋ {ŋ → ∅ N → Ṽ}
V:	{N → ŋ N → Ṽ} N → ∅	W:	{N → ŋ N → Ṽ} ŋ → ∅	C:	N → ŋ ŋ → Ṽ ŋ → ∅	B:	N → ŋ ŋ → Ṽ Ṽ → ∅	T:	{N → ŋ N → Ṽ} Ṽ → ∅
M:	N → Ṽ N → ∅ N → ŋ	Q:	N → ∅ N → Ṽ N → ŋ	J:	N → Ṽ N → ŋ ŋ → ∅	R:	{N → Ṽ N → ∅} N → ŋ	L:	N → Ṽ N → ŋ N → ∅
X:	N → Ṽ {N → ŋ Ṽ → ∅}								

[a]Log likelihoods are in decreasing order within each row and then from row to row.

ingly enough, the rate-limiting factor is not the number of token classifications but the number of rules. The token classifications are generated very quickly and evaluated quickly, given all the rule likelihoods. But to assess each rule requires considerable time. We were thus obliged to eliminate one of the variants from the data set. The one used by the fewest speakers and with the fewest tokens was [n], and so these tokens were not analyzed. It should be pointed out, however, that the distribution of [n] is similar to that of [l].

The ⟨r⟩ variants considered are [r], [l], [ř], [h], C (assimilated to the following consonant), and ∅. In discussing rule orders, we will assume that this list is ordered in terms of increasing superficiality, though this is somewhat arbitrary and has no bearing on the statistical analysis. We could assume any other underlying/surface configuration and rephrase our discussion accordingly without changing the substantive results.

Note that to represent the 813 possible rules (in the technical sense we have been using), there are only 30 different pairs of form $x \rightarrow y$ based on six variants. Assuming underlying/surface distinctions cuts this down to 15. Of course, each such pair can represent many different things, depending on the nature of the schema of which it is part. Nevertheless, it will be meaningful to discuss the results here in terms of these pairs.

The ⟨r⟩ data were coded for following segment, speech style, and sex of

TABLE 6. *Summary of* ⟨r⟩ *reduction analysis*

Best rule orders: (ordered rules only)			
r → ṙ	r → l	r → l	r → ṙ
r → l	r → h . . . h → ∅	r → ṙ	r → h . . . h → ∅
r → h . . . h → ∅	r → ṙ	r → h . . . h → ∅	r → l
r → C	r → C	r → C	r → C
	inferred:	{r → ṙ, r → l, r → h}	
		{h → ∅, r → C}	
	Worst rule orders: include	ṙ → l, ṙ → h, r → ∅, h → C	

speaker. We point out that excluded from the data are all cases where the ⟨r⟩ precedes an ⟨l⟩ – in this case, assimilation is virtually categorical and no statistical analysis is necessary.

Of the 813 rules in Table 3, those which involve 3-, 4-, or 5-way splits number fully 458. Furthermore, it is these rules that require the most computing time. Thus, we undertook to evaluate only those rule order schemata containing no subsets of unordered rules. We can hope that if the best schema does contain unordered rules, then the completely ordered schemata most similar to it will all have high likelihoods and we can still make the correct inferences.

Table 6 depicts the four best schemata (D. Sankoff, 1986). Where there is no necessary ordering between rules, this is emphasized by writing them in separate columns. From these results, as well as from the next 10 best schemata, we can infer that the most reasonable analysis is one where [ṙ], [l], and [h] are generated in an unordered way, and the assimilated and ∅ variants are then derived from [r] and [h], respectively.

We also display in Table 6 a list of those rules recurring in many of the worst (i.e., least likely) analyses. This is consistent with the fact that none of these occurs in the best analyses, and bolsters our conclusion that in ⟨r⟩ reduction, there are really four separate processes involved: lateralization, spirantization, aspiration (this is the main "weakening" path; both steps are clearly weakenings), and then assimilation.[4] This is a very different model from what has been posited through purely historical or internal phonological arguments. That it comes out so clearly in the analysis, however, is a convincing reason to accept it.

The key to the validity of an empirical analysis is its reproducibility. Although Poplack's data have not been replicated in the same community, we have been able to analyze a comparable set of data collected by Cedergren (1973) on ⟨r⟩ in Panamanian Spanish. There are a number of differences between the two data sets; in the coding scheme for her early study, Cedergren did not distinguish between the assimilated variant C and the null realization of the variable ∅. Second, not all the variants of Panamanian and Puerto Rican ⟨r⟩ are strictly comparable; the spirant [ř] has a distinctly weaker articulation in the latter than the former; the lateralized variant [l] is rare in Panama and is not included in the data set. Third, Cedergren's data include noninfinitival ⟨r⟩ as well as factor groups that code for the position in the word of the ⟨r⟩ and whether the ⟨r⟩ itself constitutes a morpheme. Nevertheless, the two data sets represent an excellent resource for the comparative study of ⟨r⟩ reduction in two communities. When we submitted the Panamanian data to a four-variant analysis (Cedergren, Rousseau, & Sankoff, 1986), an examination of the most likely rule orders clearly showed spirantization and aspiration to be separate, unordered processes. It was not clear whether deletion was still another distinct process, unordered with respect to the other two, or whether it applied to the output of the spirantization rule. It certainly did not, however, apply to the output of the aspiration rule as it did in the Puerto Rican data. This difference could be easily understood if Cedergren's ∅ tokens include large numbers of the assimilated variant, as they would if the distribution were similar to that found in the Puerto Rican variety. It could also be connected with the different nature of the spirant in the two varieties. In any case, the evidence for the independence of spirantization and aspiration in the Panamanian data provide dramatic corroboration of the results on Puerto Rican infinitival ⟨r⟩.

CONCLUSIONS

The evaluation of contextual effects on the expression of a phonological variable involves the systematic collection and analysis of many thousands of occurrences of the different variants. Under the hypothesis that these variants are derived from a common underlying form through some (unknown) series of ordered, unordered, or partially ordered rules, the statistical distribution of the occurrences across the set of possible contexts necessarily contains a great deal of information about this order (or lack of it).

Prudent use of variable rule methodology, extended in order to handle more than two variants, enables us to infer rule order, or at least to drastically reduce the myriad possibilities which exist when there are four, five, or more variables. It should be stressed that, viewed in statistical terms, the rule order inference problem reduces to a maximum likelihood hierarchical clustering of the variants according to their relative tendencies to be expressed in similar or dissimilar contexts. There is another component of rule order, which is sometimes mistakenly thought to be identical to rule order, namely, underlying/surface distinctions; but the data gathered in empirical studies of usage do not contain information about this component. Any token classification schema inferred through our methods will be consistent with a variety of different underlying/surface hypotheses.

In one sense, this is a weakness in the empirical approach to the study of rule order. This is compensated for, however, by the fact that the linguistic structures inferred through these methods retain their validity and interest outside of the particular theoretical framework inherent in the notions of phonological rule, underlying form, and rule order.

One substantive generalization that emerges from the three variables studied in this article is the relative shallowness of the reduction process for syllable-final consonants in Caribbean Spanish. The intrinsic orders s > h > ∅, n > ŋ > Ṽ > ∅ and r > ṙ > h > C > ∅, which have seemed most natural to many researchers at first view for historical reasons or by analogy to hierarchies of phonological strength, do not fare very well when confronted with performance data. Even ⟨s⟩ reduction may be equally well analyzed in terms of unordered aspiration and deletion. For the ⟨n⟩ variable, velarization and deletion are clearly independent and unordered, and although vocalization may be a second step, it probably does not operate on the output of the other two rules. In the case of ⟨r⟩, lateralization, spirantization, and aspiration are unordered, separate, processes, followed by optional assimilation of the unaffected tokens, whereas only deletion operates on the output of a preceding rule (aspiration).

NOTES

1. For computational purposes, it is convenient to work with the logarithm of the likelihood and not the likelihood itself. This practice has no effect on our evaluation of the relative merits of different token classifications.
2. In this article, we will not dwell on the interpretation of the estimates obtained for the parameters of the various rules. In some cases, these do not depend strongly on the choice of rule order schema, but in others, the interpretation is radically different in the "correct" choice than that in some of the less likely schemata.
3. As in all this work, we do not take account of other, unrelated rules of phonology that may intervene between two of these rules to affect their context and hence possibly the relative merit of one order rather than another.
4. Had we included the [n] tokens, it is likely that nasalization would also have emerged as a separate process.

REFERENCES

Cedergren, H. J. (1973). *Interplay of social and linguistic factors in Panama.* Ph.D. dissertation, Cornell University.

Cedergren, H. J., Rousseau, P., & Sankoff, D. (1986). La variabilidad de /r/ implosiva en el español de Panamá y los modelos de ordenación de reglas. In R. Núñez Cedeño, I. Páez Urdaneta, & J. Guitart (eds.), *Estudios sobre la Fonología del Español del Caribe.* Caracas: La Casa de Bello. 13–20.

Kemp, W. (1979). L'histoire récente de *ce que, qu'est-ce que* et *qu'osque* à Montréal: Trois variantes en interaction. In P. Thibault (ed.), *Le français parlé: Etudes sociolinguistiques.* Edmonton: Linguistic Research Inc. 53–74.

Poplack, S. (1979). *Function and process in a variable phonology.* Ph.D. dissertation, University of Pennsylvania.
______ (1980). Deletion and disambiguation in Puerto Rican Spanish. *Language* 56:371–86.
Poplack, S., & Walker, D. (1986). Going through (L) in Canadian French. In D. Sankoff (ed.), *Diversity and diachrony.* Amsterdam and Philadelphia: John Benjamins. 173–98.
Pupier, P., & Légaré, L. (1973). L'efacement du /l/ dans les articles définis et les clitiques en français montréalais. *Glossa* 7:63–80.
Rousseau, P., & Sankoff, D. (1989). The inference of hierarchical schemes for multinominal data. *Journal of Classification* 6:73–95.
Sankoff, D. (1986). Ordenamiento de reglas variables: /r/ implosiva en un dialecto puertorriqueño. In R. Núñez Cedeño, I. Páez Urdaneta, & J. Guitart (eds.), *Estudios sobre la Fonología del Español del Caribe.* Caracas: La Casa de Bello. 109–15.
______ (1988). Variable rules. In U. Ammon, N. Dittmar, & K. J. Mattheier (eds.), *Sociolinguistics: An international handbook of the science of language and society.* New York: de Gruyter. 984–97.
Sankoff, D., & Labov, W. (1979). On the uses of variable rules. *Language in Society* 8:189–222.
Sankoff, D., & Rousseau, P. (1982). Un diagnosticó para reglas convergentes. In O. Alba (ed.), *Español del Caribe.* Dominican Republic: Universidad Madre y Maestra. 291–300.
Sankoff, G., & Cedergren, H. J. (1971). Some results of a sociolinguistic study of Montreal French. In R. Darnell (ed.), *Linguistic diversity in Canadian society.* Edmonton: Linguistic Research Inc. 61–87.
Santerre, L., Noiseux, D., & Ostiguy, L. (1977). La chute de la consonne "l" dans les articles définis et les pronoms clitiques en français montréalais. *The Fourth LACUS Forum 1977.* 530–38.

APPENDIX

We sketch briefly here how to calculate the quantities illustrated in Table 3.

R^*_n: *the total number of rules based on at most* n *variants.*

$$R^*_n = \sum_{m=2}^{n} \binom{n}{m} R_m \ ,$$

where R_n is the number of rules based on exactly m variants. Let R^k_m be the number of rules based on m variants partitioned into k sets. For example, {2}, {4}, {1,3} is a rule counted in R^3_4. By considering the different ways a rule based on m variants can be constructed from one based on $m - 1$ variants, we have $R^2_2 = 1$, and for $m = 3, 4, \ldots$

$$R^k_m = kR^k_{m-1} + R^{k-1}_{m-1}$$

for $k = 3, 4, \ldots, m - 1$;

$$R^2_m = 2R^2_{m-1} + 1 \ ,$$

and

$$R^m_m = 1 \ .$$

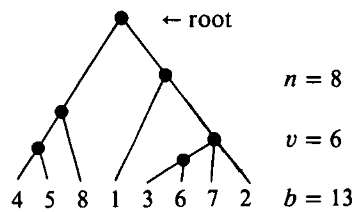
Using these recurrences we can calculate the

$$R_m = \sum_{k=2}^{m} R_m^k$$

and hence R_n^*.

T_n: the number of different token classifications based on n *variants.*

A token classification is equivalent to a rooted tree structure, for example:



Each nonterminal vertex, including the root, represents a rule and the number of branches (at least two) descending from the vertex is the number of sets in the partition constituting that rule. There are n terminals, v nonterminal vertices, and b branches in the tree, and

$$b = v + n - 1 \ .$$

If there are T_n^v different trees having v nonterminal vertices and n terminals, each such tree can be constructed from a tree having $n - 1$ terminals by suitably adding a branch to pre-existing vertex or to a new vertex placed in the middle of a pre-existing branch, thus creating two new branches out of the old one. Then $T_2^1 = 1$, and for $n = 3, 4, \ldots$

$$T_n^v = vT_{n-1}^v + [(v-1) + (n-1)]\, T_{n-1}^{v-1}$$

for $v = 2, 3, \ldots, n-2$;

$$T_n^{n-1} = (2n-3)\, T_{n-1}^{n-2} \ ;$$

and

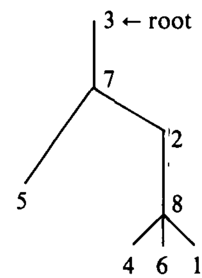
$$T_n^1 = 1 \ .$$

The multiplier of T_{n-1}^{v-1} in the recurrence exceeds by one the number of branches in trees with $n + v - 2$ vertices to account for the possibility that the root vertex represents a rule of form $\{1, 2, \ldots, n-1\}, \{n\}$. Then

$$T_n = \sum_{v=1}^{n-1} T_n^v \ .$$

N_n: *the number of different sets of underlying/surface distinctions based on* n *variants.*

A set of distinctions is equivalent to a rooted tree structure, where each of the n vertices, including terminal and nonterminal ones, represents a variant, for example:



Each nonterminal vertex may have as few as one line descending from it.

Consider any partition of n different numbers into m sets, and number the sets $K_1, K_2, \ldots, K_m$ in order of decreasing size. If two or more sets have the same size, order them according to the size of the smallest number in each of them. Let k_1, $k_2, \ldots, k_m$ be the sizes of the sets, so that

$$\sum_{i=1}^{m} k_i = n .$$

Define

$$\begin{bmatrix} & & n & \\ k_1 & k_2 & \ldots & k_m \end{bmatrix}$$

to be the number of different ways of constructing such an ordered partition.

We have n possible variants for the root of the tree, and m branches descending from it $(1 \leq m \leq n-1)$, each of which leads to another variant. Each such variant i is the root of a subtree containing a total of k_i variants, where we label the m roots according to the same convention already described. Then, it follows that $N_1 = 1$, $N_2 = 2$, and for $n = 3, 4, \ldots$

$$N_n = n \sum_{m=1}^{n-1} \sum_{(K_1, K_2, \ldots, K_m)} \begin{bmatrix} & & n-1 & \\ k_1 & k_2 & \ldots & k_m \end{bmatrix} \prod_{i=1}^{m} N_{k_i} ,$$

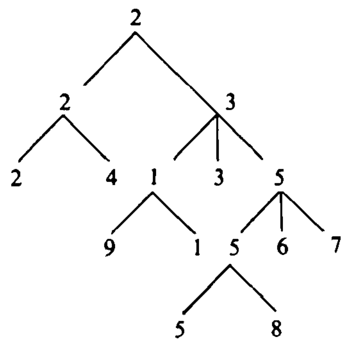
where the second summation is over all suitably numbered partitions.

It may be proved that

$$N_n = n^{n-1} .$$

M_n: *the number of different rule order schemata based on* n *variants.*

A rule order scheme is equivalent to a rooted tree structure where each vertex i must have at least two branches descending from it, exactly one of which has the same variant label as i. For example:



There are exactly n terminal vertices.

We have n possible variants for the root. One branch descending from the root labelled with variant i leads to a subtree also with root labelled i, and containing l variants in all, where $1 \leq l \leq n - 1$. Aside from i, there are $\binom{n-1}{l-1}$ ways of choosing these variants. This subtree may thus have any of $\binom{n-1}{l-1}M_l/l$ forms, because its root label is required to be i.

There is at least one other branch descending from the original root, and possibly as many as $n - l$. Using the same arguments as in the calculation of N_n, then, we have $M_1 = 1$, $M_2 = 2$, and for $n = 3, 4, \ldots$

$$M_n = n \sum_{l=1}^{n-1} \binom{n-1}{l-1} \frac{M_l}{l} \sum_{m=1}^{n-l} \sum_{(K_1, K_2, \ldots, K_m)} \begin{bmatrix} & n - l & \\ k_1 \quad k_2 & \ldots & k_m \end{bmatrix} \prod_{i=1}^{m} M_{k_i} .$$