Subject: STORAGE WebVoyage Request Get Article Copy from Storage
From: Penn Library OPAC <voyager@tx.lib.upenn.edu>
Date: Fri, 21 Sep 2007 11:01:54 -0400 (EDT)
To: storage@pobox.upenn.edu

```
Patron Info:
 Michael Friesner
215-779-0847
STORAGE WebVoyage Request: (Get Article Copy from Storage)

 Datebase: LOCAL ()
 Pick Up At: 172
 Not Needed After: 2007-09-28 00:00:00

 Comment:
 I need the following article, preferably an e-copy. Thank you very much!:
Sankoff, David (1988). Variable rules. In Ammon, U., Dittmar, N. & Mathheier, K.J.
(Eds.), Berlin Sociolinguistics: An International Handbook of the Science of Language
and Society Vol. 2. Berlin: Walter de Gruyter, 984-997.

  Bib Info:
 ---------
Other Author(s):    Ammon, Ulrich.
                    Dittmar, Norbert.
                    Mattheier, Klaus.

Title:              Sociolinguistics, an international handbook of the science
                       of language and society = Soziolinguistik, ein
                       internationales Handbuch zur Wissenschaft von Sprache
                       und Gesellschaft / edited by Ulrich Ammon, Norbert
                       Dittmar, Klaus J. Mattheier.

Primary Material:   Book

Subject(s):         Sociolinguistics--Handbooks, manuals, etc.

Publisher:          Berlin ; New York : Walter de Gruyter, 1987-1988.

Description:        2 v. ; 28 cm.

Series:             Handbücher zur Sprach- und Kommunikationswissenschaft ;
                       Bd. 3 = Handbooks of linguistics and communication
                       science

Notes:              English and German.
                    Includes bibliographies and indexes.

Location:           Storage: From RECORD page, use Place Request tab

Call Number:        P40 .S62 1987
Status:             Available, check location
```

7.13.12.1

Viereck, Wolfgang (1984) "Der Einsatz des Computers in der Amerikanisch-Englischen und Britisch-Englischen Dialektologie und Soziolinguistik", in: *Zeitschrift für Dialektologie und Linguistik* 51, 6–30.

Werner, Oswald (1982) "Microcomputers in cultural anthropology, APL programs for qualitative analysis", in: *Byte* 7, 7, 250–280.

Winograd, Terry (1983) *Language as a cognitive process, Volume 1: Syntax*, Massachusetts.

Zipf, G. (1965) *The psycho-biology of language: an introduction to dynamic philology*, Cambridge, Massachusetts.

Roeland van Hout, Tilburg (The Netherlands)

112. Variable Rules

1. Choices, Contexts and Factors

Whenever a choice among two (or more) discrete alternatives can be perceived as having been made in the course of linguistic performance, and where this choice may have been influenced by factors such as features in the phonological environment, the syntactic context, discursive function of the utterance, topic, style, interactional situation or personal or sociodemographic characteristics of the speaker or other participants, then it is appropriate to invoke the statistical notions and methods known to students of linguistic variation as *variable rules*. There are many other areas in linguistics in which statistics are useful, such as the analysis of the *continuous* variables of acoustic phonetics, but variable rule analysis (which, in present usage, does not necessarily involve 'rules' at all) pertains specifically to the probabilistic modeling and the statistical treatment of *discrete* choices and their conditioning.

One prerequisite for a variable rule analysis, then, is the perception on the part of the analyst of the existence of some sort of speaker's choice between two or more specified sounds, words or structures during performance. Furthermore, this type of analysis is only justified if the outcome of the choice process is at least sometimes unpredictable from whatever types of contextual information, such as the factors mentioned above, are known or given (so that a variable rule analysis may be rendered obsolete as more pertinent contextual information is brought to bear – a scenario often envisaged but seldom realized). Finally, the choice process must be seen to recur. Statistical inference extracts regularities and tendencies from data presumed to have a random component, which makes them appear to have less structure and more exceptions than they really do. To accomplish this, the inference procedures must be applied to some sample containing the outcomes of the choice repeated many times, usually in a variety of contexts, each context being defined as a specific configuration of conditioning factors.

Though it was first developed as a quantitative extension of generative phonological analysis and notation (Labov 1969, 715; Cedergren/Sankoff 1974, 333), it is not a prerequisite of variable rule analysis that the choice mechanism itself have any particular kind of linguistic or sociological interpretation. The statistical analysis does not depend on the origin of the variability in the data, whether it be in the grammatical generation of sentences, in processes of production and performance, in the physiology of articulation, in the conscious stylistic decisions of speakers, or even simply as an analytical construct on the part of the linguist. The linguistic *significance* of the analysis does of course depend on the nature of the choice process, but this question must be addressed prior to the formal, algorithmic, statistical procedures, e. g. in the collection and coding of the data, or in the decisions about what choice is to be studied and what is to be considered the context, and/or afterwards, in the interpretation of the results.

2. Factors and Marginals

In variable rule analysis we are given a sample of choice outcomes in various *contexts*, usually an exhaustive compilation from a corpus considered to be a sample of discourse from one or more speakers or texts. The essence of the analysis is an assessment of how the choice process is influenced by the different *factors* whose specific combinations define these contexts. While accepting that the choice cannot usually be predicted with certainty, it is still possible to ascertain what, if anything, favours a given alternative, and how strongly, and what disfavours it.

For example, consider the (fictional) data in Fig. 112.1 on plural morpheme expression in Nepean. These data indicate the frequency with which the optional suffix *-enmas* appears on the components of various lexical categories within the NP, when this NP is in subject or object position.

Objects	Subjects	Total
Adjectives:		
6/10 = 60%	10/10 = 100%	16/20 = 80%
Determiners:		
3/10 = 30%	7/10 = 70%	10/20 = 50%
Nouns:		
0/10 = 0%	4/10 = 40%	4/20 = 20%
Total:		
9/30 = 30%	21/30 = 70%	

Fig. 112.1: Data where marginal analysis is valid. Number of cases with plural morpheme expressed out of ten examples for each combination of lexical category and grammatical function.

In this example, each lexical category is a *factor*, and the set of categories comprises one factor *group*. Subject and object positions are factors, and together they comprise a second factor group. In the analysis of variance and similar statistical procedures, our factors would be called *levels* and our groups would be called factors. The set of possible choices – expression or omission of the plural morpheme – constitutes a *variable*. In regression analysis, our factors would be called *explanatory* variables and our choice variable would be the *dependent* variable.

It is of critical importance to distinguish between the roles of the (dependent) variable and the *explanatory* factors. We are interested in how likely the plurality of a given noun is to be overtly expressed versus how likely it is to be deleted, and in how probable it is that a word in object position will drop its plural marker versus how probable it is that it will keep it, but we are not interested, for present purposes, in what proportion of all potential plural markers happen to fall in object NPs or on words of various categories, and even less interested in what proportion of the words in object position are determiners and what proportion adjectives. These latter questions may well be of interest, but if, for linguistically motivated reasons, we wish to focus on plural marking as a choice process, we must treat the distribution of contexts as given, and not confuse its statistical analysis with that of the choice variable under study. Of course, some of the elements of the context may also be analyzable as the result of lingusitic choice, but this is properly done in a separate analysis. It is possible to analyze two or more linguistic choices being made simultaneously, but this is a special case and requires a slightly different statistical approach, to be discussed in Section 11. As for sociodemographic, situational or stylistic factors, there is seldom any justification to treat them as choices subject to mutual influences of phonological or syntactic performance processes – whether a speaker marks a specific occurrence of a specific noun with a plural morpheme should not be analyzed as having an immediate effect on the speaker's age or sex, the interlocutors present or the degree of formality adopted. If any such factor is to be analyzed as a choice, this must be considered as having occurred prior to performance choices, in some temporal, generative or hierarchical sense.

To analyze the data in Fig. 112.1, it might seem that it would suffice to compare the overall percentage of occurrences where the plural marker was expressed rather than omitted on nouns, on adjectives and on determiners, and similarly to compare overall rates of expression in subject and object positions. No more statistical analysis is necessary to confirm that subject position and adjectives favour expression, while objects and nouns favour omission. The relative sizes of the marginal totals for nouns, adjectives and determiners are faithfully reflected both within the subject data alone, and within the object data alone. Similarly, the marginals for subject and object are paralleled within each of the three lexical categories. There is no question of confusing the relative effects of lexical category with that of grammatical function.

This type of analysis, called the *comparison of marginals*, is not always so clearly correct. In fact the data set in Fig. 112.1 has been artificially constructed so that an examination of the marginals is all that is needed. Note that each context is represented by exactly ten cases. If this kind of even distribution is not found, the marginals can give a very distorted impression of the effects of the factors. For example, the percentages in each context in Fig. 112.2 are the same as in Fig. 112.1 but the marginals do not give the correct picture at all.

Objects	Subjects	Total
Adjectives: 42/70 = 60%	10/10 = 100%	52/80 = 65%
Determiners: 3/10 = 30%	49/70 = 70%	52/80 = 65%
Nouns: 0/10 = 0%	76/190 = 40%	76/200 = 38%
Total: 45/90 = 50%	135/270 = 50%	

Fig. 112.2: Data where marginal analysis is grossly misleading. Number of cases with plural morpheme expressed out of all examples for each combination of lexical category and grammatical function

The marginals here do not reflect the fact evident in both subject and object positions that adjectives favour expression more than determiners do. Nor do the marginals indicate the clear tendency within all three lexical categories for subjects to favour expression more than objects.

The type of uneven distribution illustrated in Fig. 112.2 is characteristic of corpus-based sociolinguistic research, where the number of occurrences of each context depends on its relative frequency in discourse. Hence the number of cases per context is highly variable and many combinations of factors may not occur at all. This contrasts, for example, with psycholinguistic studies, which are usually based on testing methodology rather than on corpus work, and which can thus assure the same number of examples in all contexts. It is a tenet of corpus-based sociolinguistics that data analysis should make use of the naturally occurring frequencies of the contexts, even if these are not statistically examined in the same way or at the same time as the choice variable. It is clear in Fig. 112.2 that the lexical and functional groups of factors themselves have a particular relationship in Ñepean — adjectives are heavily utilized in object position, whereas subjects almost all consist of nouns plus the occasional determiner. When this type of *dependence* among factor groups obtains, the examination of the marginals may give highly misleading results.

3. Interaction of Factor Effects

Even when the data are uniformly distributed among the contexts, we still cannot always be confident in an analysis of the marginals only. Fig. 112.3 exemplifies data for which this leads to serious error.

Objects	Subjects	Total
Adjectives: 9/10 = 90%	7/10 = 70%	16/20 = 80%
Determiners: 0/10 = 0%	10/10 = 100%	10/20 = 50%
Nouns: 0/10 = 0%	4/10 = 40%	4/20 = 20%
Total: 9/30 = 30%	21/30 = 70%	

Fig. 112.3: Data illustrating independent factors which interact

Here, although the marginals are identical to those of Fig. 112.1 and although the factors are completely independent, subject position favours plural morpheme expression more than object position does only on nouns and determiners. By the same token, adjectives favour expression more than the other categories only in object position. Thus the marginals do not reflect the contextual patterns of expresion versus omission — we say that the factors have *non-independent effects*, or that they *interact*, even though the factor groups are themselves perfectly independent in terms of their cross-cutting distribution. Note that our terminology is that of regression analysis, and not that of contingency tables (e. g. loglinear analysis) where dependence and interaction are not distinguished.

4. The Null Hypotheses

A primary goal of variable rule analysis, then, is to allocate contextual effects on the choice process among the different cross-cutting factors which make up the contexts, in a way which is able to correct both for dependent factors and for interacting factors. This is a well-known type of problem in statistical analysis: it may be formalized in a number of ways, and there may be different approaches to its solution within each formulation. We will briefly sketch some of the possibilities and explain why the variable rule approach has evolved as most appropriate for sociolinguistic data.

Statistical methodology is generally elaborated with a view to distinguishing bona fide differences and trends from accidental data patterns due to statistical fluctuation (or noise, or random error). The hypothesis that there are no genuine effects in the data is called the *null hypothesis*, and it is this hypothesis which must be falsified, on the basis of a set of data, if we are to convince ourselves that a real effect exists.

For the type of data which interests us, namely choice frequencies in contexts made up of cross-cutting factors, the null hypothesis is that none of the factors examined has any systematic effect on the choice process, and that any differences in choice outcome proportions among the various contexts must be attributed to statistical fluctuation. If we can prove that random processes alone are unlikely to have resulted in the pattern of proportions observed, we may be able to attribute this pattern to the effect of one or more of the factors. There are two ways of looking for systematic deviations from randomness. One way is to see if there is any single factor or group which seems to contradict the null hypothesis. If there is, we take into account how much of the patterning in the data it explains, and then we search for any further systematicity in the remaining variability to see if it may be explained by a second factor, and so on. After the search for these factors, or during it, we can also check whether or not there are any interactions among them. We will discuss this approach more fully in Section 9.

A contrasting approach, which we present in Section 11, is to try to account for all possible interactions among groups of factors at the same time as we establish individual factor effects. This kind of procedure, which can be found in some commercial computing packages, becomes unwieldy when the number of factor groups is more than three or four, as is often the case in linguistics.

5. Models and Link Functions

To separate the effects of different contextual factors, it is important to have an idea of how these effects combine in influencing the choice process in a given context. The simplest way of combining effects is additive – if two factors, each of which is thought to have some quantitative effect, are both present when the choice is made, then their combined effect should be the sum of their individual contributions. Furthermore, the individual contribution of a factor should remain invariant across all contexts containing that factor. For example, if the factor effects are as in Fig. 112.4, then the percentage predicted for each context in Fig. 112.1 by summing the effects of the two factors which define that context are exactly as found empirically.

Adj.	Determ.	Noun	Object	Subject
80%	50%	20%	−20%	20%

Fig. 112.4: Effects of individual factors in Fig. 112.1.

In this case we say that the additive model fits the data exactly. Though the figures in Fig. 112.4 bear an obvious relationship to the marginals in Fig. 112.1, this is not always so. For example, the effects in Fig.112.4 also fit perfectly the data in Fig. 112.2, though there is no relationship to the marginals in this case.

It is important to note that not all data can be well fit by an additive model for combining effects. For example, no additive model accounts well for the data in Fig. 112.3.

The additive model has the virtue of simplicity and is often quite appropriate when there are only two or three groups of cross-cutting factors. It is the basis for statistical procedures such as the analysis of variance. One weakness it does have, however, especially when there are many factor groups as is so often the case in sociolinguistics, is that it becomes impossible to find factor effects which enable the model to fit the data well in most contexts, without having it predict, for certain other contexts, percentages which are greater than 100% or less than 0%. Consider Fig. 112.5.

	Following Vowel		Following Consonant	
	Obj.	Subj.	Obj.	Subj.
Adjectives:	8/10	10/10	4/10	10/10
Determiners:	5/10	9/10	1/10	5/10
Nouns:	0/10	6/10	0/10	2/10

Totals	
Adjectives:	32/40 = 80%
Determiners:	20/40 = 50%
Nouns:	8/40 = 20%
Objects:	18/60 = 30%
Subjects:	42/60 = 70%
Following Vowel:	38/60 = 63.3%
Following Consonant:	22/60 = 36.7%

Fig. 112.5: Data where additive model predicts less than 0% and more than 100%

The additive model which fits these data best is summarized in Fig. 112.6.

Adjectives:	80%
Determiners:	50%
Nouns:	20%
Objects:	−20%
Subjects:	20%
Following Vowel:	13%
Following Consonant:	−13%

Fig. 112.6: Factor effects for data in Fig. 112.5 under the additive model

By adding the effects of one factor from each of the three groups, lexical category, grammatical function and following phonological segment, we arrive at a *prediction* of the rate of expression of the morpheme *-en-mas* in the context defined by this configuration of factors. In Fig. 112.7 we compare predicted and observed percentages. One prediction is found to be greater than 100% and one is less than 0%.

By themselves, such 'impossible' predictions may seem like a minor annoyance, but when it comes to comparing data sets, evaluating competing analyses, testing for statistical significance and interpreting the effects of certain factors, this property of the additive model becomes a major problem.

	Following Vowel		Following Consonant	
	Obj.	Subj.	Obj.	Subj.
Adjectives:	73	113	47	87
	(80)	(100)	(40)	(100)
Determiners:	43	83	17	57
	(50)	(90)	(10)	(50)
Nouns:	13	53	−13	27
	(0)	(60)	(0)	(20)

Fig. 112.7: Percentages predicted under the additive model compared to observed (in parentheses) percentages for the data in Fig. 112.5

When this type of problem arises in statistical modeling, the solution is to use a model where the sum of the factor effects is not the predicted percentage of a given choice, but some quantity related to this percentage by a *link function*: this function is such that it can take on any value without the risk that the corresponding percentage will be less than 0 or more than 100. In variable rule analysis, the link function is usually the *logit* of the percentage:

$$\text{logarithm}\left(\frac{\text{percentage}}{100 - \text{percentage}}\right) = \text{sum of factor effects}$$

or equivalently since the probability p of a choice is just the predicted percentage of that choice divided by 100:

$$\log\left(\frac{p}{1-p}\right) = \text{sum of factor effects}$$

Fig. 112.8 shows the factor effects under the logit-additive model (also called logistic-linear or logistic regression) for the data in Fig. 112.5.

Adjectives:	2.2
Determiners:	0
Nouns:	−2.2
Objects:	−1.5
Subjects:	1.5
Following Vowel:	1.0
Following Consonant:	−1.0

Fig. 112.8: Factor effects on Nepean plural expression, calculated from data in Fig. 112.5 under the logit-additive model. Cf. Sections 7 and 8 for method

Fig. 112.9 compares observed and predicted proportions under this model. Note that the logit-additive model generally fits the data as well as or better than the additive model. (cf. Fig. 112.7).

The logit has two properties which make it superior to other functions which have been suggested. First, whatever the sum of the factor effects, zero, any positive number or any negative number, the predicted percentage always lies between 0 and 100, as in Fig. 112.9 (and p lies between 0 and 1). This condition does not automatically hold (though it can be imposed) for certain other link functions such as log p or -log (1-p) used in early variable rule work (Cedergren/Sankoff 1974, 333).

Following Vowel Obj.	Subj.	Following Consonant Obj.	Subj.
Adjectives:			
85	99	43	94
(80)	(100)	(40)	(100)
Determiners:			
58	92	17	62
(50)	(90)	(10)	(50)
Nouns:			
6	57	1	25
(0)	(60)	(0)	(20)

Fig. 112.9: Percentages predicted under the logit-additive model compared to observed (in parentheses) percentages for the data in Fig. 112.5

Second, the logit is symmetrical in a very important sense with respect to binary choices (A or ~A, e. g. expression versus omission). If we analyze the data in terms of ~A instead of in terms of A (e. g. in terms of deletion instead of expression) the model has exactly the same form; all that changes is that p becomes the probability of ~A instead of A, and the sign (positive or negative) of each factor effect is reversed, so that a positive effect of a factor on the choice of A is the same as a negative effect on the choice of ~A, and vice-versa. This symmetry can be seen through the comparison of Fig. 112.8 and 112.10.

Using link functions such as log p or -log (1-p), on the other hand, the relationship between the models for the choice of A and for the choice of ~A is much less natural. A positive factor effect in favour of A is not an identical negative effect in favour of ~A, but obeys a much more complicated relationship. A further manifestation of the symmetry of the logit is that when p is close to 1, changes in factor effects induce the same size changes in p as when p is close to zero. The log p link function, on the other hand, makes the probability very sensitive to effect changes when p is close to 1 and very insensitive when p is close to 0. The same holds in the reverse direction for the -log (1-p) model. Unless we have some good reason for carrying out the analysis in terms of one of A or ~A but not the other (and this is seldom the case with sociolinguistic data), this asymmetry is an arbitrary and unjustified assumption in these models. Except where otherwise stated, then, the adoption of the logit link function will be implicit in our discussion of variable rules.

Adjectives:	−2.2
Determiners:	0
Nouns:	2.2
Objects:	1.5
Subjects:	−1.5
Following Vowel:	−1.0
Following Consonant:	1.0

Fig. 112.10: Factor effects on Nepean plural *omission*, calculated from data in Fig. 112.5 under the logit-additive model

6. Constraints, Input Probability and the Corrected Mean

It is a property of models based on a sum of factor effects that many different sets of values for these effects will predict the same context proportions. It is easy to see that the figures in Fig. 112.4 could be replaced by those in Fig. 112.11 without any change in the predicted percentages.

Adj.	Determ.	Noun	Object	Subject
60%	30%	0%	0%	40%

Fig. 112.11: Effects of individual factors in Fig. 112.1

This may seem an undesirable property, but it is merely a consequence of the fact that it is the *comparison* of the effects of any two factors in a factor group (as measured by their difference) which is important, and not their individual values. Thus we may add an arbitrary amount to all the factors in a group, and subtract the same amount from all the factors in another group, without changing the comparative size of any two factors within a factor group, nor the percentages predicted by the additive model (cf. Fig. 112.4 and 112.11). The same property holds for any model based on the additivity of the factor effects, including the logit-additive model.

This non-uniqueness is more of a nuisance than a conceptual defect, but in order to be able to compare different analyses, it is desirable to remove it from the model. As is done in the analysis of variance and other statistical methods, given any set of factor effects, we simply subtract the average value of each factor group from each of its factors, to give a *centred* group of factors. To compensate, we introduce a new term m in the formula:

$$\log\left(\frac{p}{1-p}\right) = m + \text{sum of factor effects}$$

m is just the sum of all the averages which were subtracted from the different factor groups. m is variously called the *input, input probability, mean effect* or *corrected mean.* Under the constraint that each factor group is centred, there is normally only one set of factor effects which predicts any particular set of context percentages based on the new formula.

7. Fitting Models to Data

Fig. 112.1 and 112.2 presented contrived data which could be fitted exactly by an additive model, while the latter could not account for Fig. 112.3 and 112.5 quite as well. In the former cases there could be no doubt about what values of the factor effects are appropriate, but when there is no set of values which enables the model to fit perfectly, which is almost always the case, the question arises of finding a set of values which best account for the observed data.

In statistics, how well a model with given factor effects fits a data set is measured by one of several criteria (e. g. the sum of squares of differences between observed and predicted values, or the sum of squares divided by the predicted values). One of the most fundamental criteria, and the one used in variable rule theory, is the *likelihood* criterion. For sociolinguistic data, this criterion is preferable both to the sum of squares familiar from the analysis of variance, multiple regression and other widely used statistical analyses, and to the sum of normalized squared differences used in χ^2 tests on contingency tables, because of the extreme distributional imbalances, including many empty cells, in corpus-based data, which tend to lead to incorrect results under the latter methods, but are not a problem for the likelihood method.

The likelihood of a set of factor effects in the case of binary choice is the product, over all contexts in the data table, of terms of form $Kp^r\ (1-p)^s$, where r is the number of occurrences of A in the context and s is the number of $\sim$A. The probability p is calculated from the factor effects using the logit-additive formula of Section 5 or whatever other model is being used, and the constant K does not depend on the factor effects at all, and so may be ignored when comparing different sets of factor values. The likelihood measure, whose mathematical derivation and properties may be found in statistics texts, indicates how likely a particular set of data is to have been generated by the model with the given values for the factor effects. Different sets of factor effects will have different likelihoods for the same set of data. Thus the principle of *maximum likelihood* suggests that a way of estimating the factor effects in a model to best account for the data is to choose that set of values which is the *most* likely to have generated the data.

For example, the values in Fig. 112.4 are maximum likelihood estimates of the factor effects, using an additive model, for the data in Fig. 112.1. The values in Fig. 112.8 and 112.10 are maximum likelihood estimates, using a logit-additive model, for the data in Fig. 112.5.

8. The Variable Rule Programs

It is one thing to state a criterion that estimates should satisfy, and quite another to actually find the values which satisfy it. Though, given any data set, the maximum likelihood estimates for the factor effects in the logit-additive model are precise, well-defined numbers, there is no simple formula for calculating them. They can be found as accurately as is desired, however, through some method of successive approximation, usually requiring a computer.

Maximum likelihood estimation in the context of the logit-additive model is not unique to linguistics, and a number of commercial statistical computing packages can carry out this analyisis, usually under the name *logistic regression.* In linguistics, however, much use has been made of *variable rule programs*, which are specifically set up to receive the type of data generated in studies of language variation, and which calculate the results in a form most useful in these studies.

Two programs have been widely circulated (non-commercially) in various versions: VARBRUL 2S and VARBRUL 3. The first of these is analogous to a stepwise multiple regression, and will be discussed in Section 8. This program was originally written by the author and was improved by researchers at the University of Pennsylvania, most recently by S. Pintzuk, who has made available a powerful and well-documented version for microcomputers. VARBRUL 3 was developed by P. Rousseau of the Université du Québec à Montréal. It is a highly versatile package capable of handling a large number of factors in carrying out the analyses we will discuss in Sections 10, 12, and 14. It is available for implementation on a number of different mainframe computers.

9. Significance and Stepwise Regression

Given any collection of choice data, categorized by a number of cross-cutting factor groups, there will generally exist maximum likelihood estimates for all the factor effects (see Section 12 for exceptions). Even if many of the factors have no real impact on the choice process, so that their contribution to the logit-additive model should logically be zero, the values estimated from the data will almost always be different from zero, solely because of sampling fluctuations in the data set. How, then, can we distinguish those factors which have a genuine effect from those whose apparent contribution is an artifact of the particular data sample? This cannot be answered on the basis of the relative size of the effects, since large effects based on few data are often illusory, while some very real effects just happen to be small in magnitude.

The pertinent criterion of validity in statistics is the *level of significance*. If we carry out a first variable rule analysis without taking into account a certain factor or factor group, and then repeat the analysis this time including that factor or group, the likelihood will always increase (or possibly stay the same) from the first to the second analysis, simply because with more 'parameters' to estimate, the fit of the model to the data cannot possibly be worse. Indeed, were the factor or group in question completely irrelevant to the choice process – the null hypothesis – statistical theory could predict roughly how much the likelihood would increase in the second analysis. Now, if the observed increase is much greater than predicted under the null hypothesis, we can conclude that the factor or group is *not* irrelevant, that it does have a statistically significant effect, be it large or small, on the choice process. The actual measurement of the level of significance is carried out by calculating:

$$2 \log \left(\frac{\text{likelihood of 2nd analysis}}{\text{likelihood of 1st analysis}} \right)$$

and then looking up the resulting value in a χ^2 table with the number of degrees of freedom equal to the number of independent factors which are added in the second analysis. Note that because of the constraints discussed in Section 6, the number of independent factors in a group is one less than the total number of factors in the group.

In the VARBRUL 2S program this is all carried out automatically, in a stepwise fashion, in order to find the configuration of factor groups most appropriate to a given data set. As discussed in Section 4, the first step is to find a single significant factor group. This is done by testing all factor groups to see which one increases the likelihood most significantly when added to an analysis based only on:

$$\log \left(\frac{p}{1-p} \right) = m$$

If no group is significant, say at the 5% level, we can conclude that none of the factors considered has any influence on the choice process. Otherwise the program retains the most significant group and tries to add a second group which will again increase the likelihood as significantly as possible. It continues in this way until no further additions of factor groups contribute statistical significance to the model. We call the collection of groups thus incorporated in the model the *step-up* solution.

The *step-down* solution is based on the same principle, but in reverse. The program starts by calculating the likelihood of the model using all the initial factor groups. It then tries to discard the group whose loss least significantly reduces the likelihood, by the same χ^2 test as above. If all groups are significant, then we conclude that they all affect the choice process. If, on the other hand, one can be discarded, the program searches for another non-significant factor among those that remain, and so on.

Ideally, the step-down analysis stops discarding groups when it is left with just the set of groups that were added in the step-up analysis. In this case, we can be fairly sure that this is the optimal group of factors. Occasionally, the two analyses do not coincide in this way. In this case, the groups which were neither added by the step-up nor discarded by the step-down, and those that were both added and discarded, remain of uncertain status.

VARBRUL 2S assesses whole factor groups at a time. The search for pertinent factors or factor distinctions *within* a factor group can be carried out in an analogous stepwise fashion, though no automatic procedure has been implemented, so that separate runs of the program are necessary for all the trial analyses (cf. Sankoff/Labov 1979, 199).

10. 'Different Grammars'

It is often of interest in sociolinguistics to compare the performance of different speakers. One way of doing this is to consider the different speakers in the corpus as separate factors. Versions of VARBRUL 2S can handle a factor group representing 20 to 30 speakers. VARBRUL 3 will accept 100 to 150.

Implicit in this approach is that speakers all have the 'same grammar' with respect to the choice process under study. In other words, speakers differ only by their overall tendency to choose A instead of ~A, as reflected in the difference among their individual factor effects. Another way of looking at this is through the implicit assumption that there is no interaction between speaker factors and other factors representing aspects of the linguistic structure within which the choice is embedded. In Section 11, we will discuss various approaches to detecting and dealing with interaction in general, but here we will concentrate on the particular case of heterogeneity among speakers.

VARBRUL 3 has the capacity of searching for ways of dividing up the sample of speakers into some specified number of subgroups, and calculating entirely separate sets of linguistic factor effects within each subgroup. Moreover, it is the program itself which constructs the subgroups and assigns speakers to them, using the principle of maximum likelihood to find the most likely division of the sample. The program can also assess the results to see if allowing each subgroup to have its own 'grammar' is worthwhile, in terms of a significant increase in the likelihood of the entire analysis.

This capability of VARBRUL 3 can only be used fruitfully on well-structured data sets; usually many speakers, several factor groups and much data distributed among a large number of contexts. Rousseau/Sankoff (1978 a, 97) give a number of examples.

11. Dealing with Interaction

The previous section discussed a method for analyzing interaction between factors representing speakers and factors representing linguistic influences on the choice process. Empirical studies of speech communities which are relatively homogeneous socially have tended not to find such interaction; with some exceptions, age and sex differentiation in phonology, morphology and syntax is generally quantitative rather than qualitative, and it is only by incorporating data from ethnically distinct groups or geographically separate communities in a study that we can expect to find 'different grammars'. Nor is interaction common among properly linguistic factor groups. That interaction which is found is often a result of infelicitous coding definitions for factors or inappropriate construction of factor groups, so that the discovery of interaction is frequently the impetus for the reformulation of the linguistic analysis of the choice process.

Interaction among extralinguistic factors, however, is quite usual.

The systematic detection of interaction in variable rule analysis can theoretically be carried out in the same way as it is in some multiple regression programs. For each pair of cross-cutting *factors* A and B, we introduce a new factor *group* with two factors, one representing the co-occurrence of A and B in a context, the other representing the absence of one or both. If this new group is statistically significant, in a VARBRUL 2S analysis, for example, then we may conclude that A and B interact. Since variable rule analyses characteristically involve ten, twenty or more factors, however, dozens or even hundreds of new interaction factor groups must be added for a truly systematic search of pairwise interactions. This, as was mentioned in Section 4, is too unwieldy, even without considering three-way, four-way, etc. interactions among factors. The technique of adding interaction factor groups to the analysis remains a useful way of *measuring* suspected interactions as well as *testing* them for significance, but it is not a feasible way of *detecting* them systematically when there are many cross-cutting factors.

In the step-up part of VARBRUL 2S, as each significant factor group is added to the analysis, the estimated factor effects of the previously incorporated factor groups will tend to change to some extent. When these changes are small, e. g. if they do not affect the way in which factor effects are ordered by size, then we may generally attribute them to sampling fluctuation. If one or more of the changes are large, then we may suspect that the new factor group and the one(s) subject to this change either are non-independent (as in Section 2) or they interact (Section 3), or both. Non-independence may be verified by simply examining the distribution of contexts, and it represents no defect in the analysis, but interaction should be tested for and incorporated in the model using interaction factor groups, or the technique described in the following paragraph.

It is often felt, especially in phonology, that binary factor groups constitute a more elegant way of representing linguistic distinctions than groups containing three or more factors. Indeed, quantitatively speaking, using two binary groups in place of one group with four factors is a stronger, more economical representation of the data, involving three independent numbers instead of four. (The binary model contains *m* and one independent factor per group, while the four-factor analysis involves *m* and three independent factor effects, as explained in Section 9.) When the strong quantitative claim implicit in the binary decomposition of choice contexts is not warranted, this is manifested as interaction in a variable rule analysis. Instead of adding interaction factor groups, which would be an unnecessary complication in this situation, we can simply use one four-factor group instead of two binary groups. In general, we can remove the interaction from two interacting groups, one with M factors and one with N factors, by combining them into a single group containing MN factors. This should only be done when there is clear interaction, since it results in a weaker model of the choice process being studied.

A further comment pertains to the situation where there is interaction only among sociodemographic factors — a common situation. Here it is often preferable to replace all these factors by a single 'speaker' factor group (cf. Section 10). The individual speaker effects can then be used as input for traditional statistical analyses, such as multiple regression, which can handle interaction efficiently.

With the availability of *loglinear analysis* in commercial computing packages, a number of studies have made use of this method to analyze linguistic variation data (e. g. Hout 1984, 39). This type of analysis is a general procedure for analyzing occurrences (or *count data*) classified by a number of cross-cutting dimensions. It is not required that any of these dimensions be specified to be dependent variables nor that others be explanatory variables. The automated treatment of interaction in these programs is often cited as a particularly attractive feature. It should be noted that the logit-additive model, with or without terms for interactive factor effects, is in effect a special case of the loglinear model. It is this special version of the loglinear model which should be used when there is one dependent variable and the rest are possible explanatory variables (Bishop/Fienberg/Holland 1975, 357), as is the case with linguistic variation data (cf. Section 2). *The more general loglinear model is inappropriate for conditioned choice data of this sort.* The mathematical reasons for this are most clearly presented by Kalbfleisch (1984, 139), and stem from the fact that the analysis treats all the factor groups in the same way as the choice variable. Thus the program tries to fit the dependence of all the factor groups on each other. This is not only conceptually absurd in many cases (cf. Section 2), but it is also statistically incorrect when any of the extralinguistic factors was not sampled in an *unstratified* way from the speech community, text base or other population under study. Equally important, the estimation of dependence among the factor groups means that proportionately less consideration is given to precisely estimating the effects of these factors and their interactions on the choice variable (Mantel 1979, 93). Because of these and other technical defects of the general approach, loglinear modeling is not recommended for conditioned choice data, except for (i) the specific version which is equivalent to the logit-additive model, and (ii) other special versions designed to handle two or more designated factors as simultaneous choice processes.

12. Implicational Scales

When only two factor groups are being examined, often 'speakers' versus 'contexts', much can be learned just by judiciously rearranging the rows and columns of the data table. Thus Fig. 112.12 may not at first glance contain any striking patterns, but Fig. 112.13 which contains exactly the same data, reveals a high degree of organization.

Context				
	U	V	W	X
Speaker				
4	0/11	0/3	7/7	0/1
2	10/10	0/3	3/3	1/1
1	1/1	5/5	4/4	2/2
3	6/6	0/1	1/1	0/9

Fig. 112.12: Choice of A instead of ~A by four speakers in four contexts

The pattern in Fig. 112.13 is called an implicational scale, because the fact that a speaker chooses A in a particular environment implies that that speaker, as well as all speakers higher in the table, choose A in that context and in all contexts further to the right in the table. Similarly, the fact that a speaker chooses ~A in a particular environment implies that that speaker, as well as all speakers lower in the table, choose ~A in that context and in all contexts further to the left in the table.

Context				
	V	X	U	W
Speaker				
3	5/5	2/2	1/1	4/4
2	0/3	1/1	10/10	3/3
4	0/1	0/9	6/6	1/1
1	0/3	0/1	0/11	7/7

Fig. 112.13: Data from Fig. 112.12, with rows and columns re-ordered to reveal implicational scale

Early discussions of variation (e.g. Bickerton 1973, 29 ff) considered implicational scales and variable rules as competing models for the analysis of choice data. Indeed, to the extent that speakers and contexts can be ordered in an implicational pattern, and to the extent that in all contexts a speaker shows either 100% A or 100% ~A, this does constitute a stronger statement about the data than does a statistical analysis.

Few substantial data sets conform to the perfect scale model, however. Even if a speaker uses 100% A or 100% ~A in most contexts, there is usually at least one 'transitional' context, where A and ~A both occur. Furthermore, no matter how the rows and columns are permuted there are usually a number of entries in the data table which violate the scaling property enunciated above. Some linguists have counted the percentage of such scaling errors out of the total number of entries and have considered the scale satisfactory if these are less than some threshold, say 5% or 10%. This type of test is totally unfounded, however, unless there is some statistical model for predicting the distribution of errors when the scaling hypothesis is true and/or when it is not true (cf. Sankoff/Rousseau 1974, 3).

One of the reasons variable rule analysis seemed incompatible with implicational scales was the property discussed in Section 5 whereby the logit-additive model predicts only percentages *between* 0% and 100%. With finite values for the factor effects, it cannot predict *exactly* 0% or *exactly* 100%, since these require an infinite value for log $(p/1\text{-}p)$. Implicational scales, on the other hand, are made up entirely (or almost) of entries containing 0% or 100%.

It had been noted that some data sets submitted to variable rule programs caused 'singular estimates' of the factor effects, namely anomalously large positive or negative numbers. Rousseau/Sankoff (1978 b, 603) discovered a mathematical characterization of such data sets, which essentially involve infinite factor effects. These turned out to be the class of data sets which can form perfect (or almost) implicational scales! Furthermore, despite the fact that finite effects cannot be estimated for all factors in these cases, predicted percentages in all contexts can be calculated with no difficulty within the maximum likelihood framework, and these *can* take on the values 0% and 100%. The discovery that implicational scales are a special kind of variable rule analysis led to a procedure (Sankoff/Rousseau 1980, 7; 1981, 257) for discarding *most likely scaling errors* from a data set in the search for an underlying perfect scale. This technique is one of the analyses incorporated in VARBRUL 3.

13. Rare Variants

In syntax, variation theory sometimes encounters a situation where one of the two variants is extremely rare, and may not even occur once in the entire corpus at hand. This variant may well represent an important phenomenon, however, such as an incipient change or a necessary condition of some other rare contextual feature. Examples of the rare variant may only be collected through what is effectively participant observation, writing down sentences as they are heard in daily conversation, on the radio or television, or read in the written media.

We then basically have two corpora, the conventional one containing no examples of the rare variant, but sufficient examples of the usual variant distributed among a sample of contexts, and the new corpus containing exclusively examples of the rare variant, presumably in a representative sample of the contexts in which it occurs.

How can we carry out a variable rule analysis if the occurrences are not drawn from the same sample of contexts? This problem has been studied in statistics (e. g. Bishop 1969, 119) and the somewhat surprising solution is that the analysis can be carried out as normal on the combined data set, as if they all originated in the same corpus. The only difference is that the estimate of m has no meaningful interpretation, since this would depend strongly on the (unknown) total amount of conversation from which the rare variant examples were extracted. The factor effect estimates, however, have their normal intepretation and are not affected by the dual origin of the data.

14. Multiple Variants and Rule Order

In the preceding sections, we have only discussed choice variables which are dichotomous, although the factor groups could contain any number of factors. There are many cases, however, where a linguistic choice may be perceived as involving three or more alternatives simultaneously.

Moreover, every topic we have discussed in connection with dichotomous variables — marginals, link functions, loglinear analyses, implicational scales, rare variants, etc. — can all be viewed in the more general context of multiple variants. Generalization of the mathematical treatment to this context offers no conceptual difficulty, only a number of complications in the formulae, and considerable programming work.

For example, consider the case of three variants A, B and C. The logit-additive model becomes:

$$\log\left(\frac{p(\mathrm{A})}{p(\mathrm{B})}\right) = m(\mathrm{A}) + \text{sum of A-effects}$$

$$\log\left(\frac{p(\mathrm{B})}{p(\mathrm{C})}\right) = m(\mathrm{B}) + \text{sum of B-effects}$$

$$\log\left(\frac{p(\mathrm{C})}{p(\mathrm{A})}\right) = m(\mathrm{C}) + \text{sum of C-effects}$$

where the context probabilities of the three variants are $p(\mathrm{A})$, $p(\mathrm{B})$ and $p(\mathrm{C})$, and three sets of inputs and other effects are postulated. Actually, the third equation is equivalent to the sum of the other two, and hence is redundant. In general, if there are n variants, then $n - 1$ equations suffice to define the model (even when $n = 2$).

VARBRUL 3 is equipped to estimate the factor effects in the logit-additive model when there are more than two variants.

When there are multiple variants, it is not always clear whether they should be considered as having been generated simultaneously, or whether certain distinctions among the variants are decided before others. This part of the classical problem of rule ordering can be studied through variable rule analysis. The other part of the problem, the question of what forms underlie what others, cannot be approached through statistical analysis of conditioned choice data.

For example, there are actually four surface forms of the plural morpheme in Nepean; besides the full form *-enmas* (variant 1) and the null form (variant 4), there are two reduced forms *-ems* (variant 2) and *-as* (variant 3). Mathematically speaking, there are 184 ways of generating these by different rule order schemes from a common underlying form, though many of these may be implausible. Two schemes are depicted in Fig. 112.14.

$$\text{(a)}\quad \begin{array}{ccc} 1 & \rightarrow & 2 \\ 2 & \rightarrow & 3 \\ 3 & \rightarrow & 4 \end{array} \qquad\qquad \text{(b)}\quad \begin{array}{ccc} 1 \rightarrow & \rightarrow & 2 \\ & \rightarrow & 3 \\ 2 & \rightarrow & 4 \end{array}$$

Fig. 112.14: Rule order schemes: (a) some occurrences of underlying variant 1 rewritten as 2, then some 2 rewritten as 3, then some 3 rewritten as 4. (b) some 1 rewritten as either 2 or 3, then some 2 rewritten as 4

The first rule in scheme (a) can be analyzed by a two-variant variable rule, since we know how many times in each context the choice was made to rewrite (the number of surface occurrences of variants 2, 3 and 4) and how many times it was not (the number of occurrences of variant 1). Similarly, the choices represented by the second rule can be analyzed using the number of variant 2 versus the number of 3 and 4. The third rule is based simply on the number of 3 versus the number of 4. In scheme (b), the two-pronged rule represents a simultaneous three-way choice between variants 1, 2 and 3, and can be analyzed using the number of variant 1 versus (2 plus 4) versus 3.

The likelihood of each of the 184 schemes can be calculated as the product of the maximum likelihoods of the individual choice processes represented by its rules. We can then invoke the principle of maximum likelihood to infer the best scheme. Analysis of data not presented here shows that (b) in Fig. 112.14 is the most likely scheme.

There are, however, five other schemes, such as the two in Fig. 112.15, which have exactly the same likelihood, since they imply exactly the same choice processes.

$$\begin{array}{lcl} 3 \to & \begin{smallmatrix} \to \\ \to \end{smallmatrix} & \begin{smallmatrix} 2 \\ 1 \end{smallmatrix} \\ 2 & \to & 4 \end{array} \qquad\qquad \begin{array}{lcl} 4 \to & \begin{smallmatrix} \to \\ \to \end{smallmatrix} & \begin{smallmatrix} 1 \\ 3 \end{smallmatrix} \\ 4 & \to & 2 \end{array}$$

Fig. 112.15: Two schemes with the same likelihood as scheme (b) in Fig. 112.14

They differ only according to which forms underlie which others. Thus only the order in which the distinctions among subsets of variants are drawn can be studied statistically, and not the question of underlying forms.

Examples of this kind of analysis are reported in Sankoff (1986, 109) and Cedergren/Rousseau/Sankoff (1986, 13).

15. Issues and Controversies

In this article, we have not discussed questions which are epistemologically anterior to the recognition of linguistic choice processes among discrete alternatives, in predetermined contexts defined by a number of cross-cutting factors. Some of these questions, including the role of variation studies in a socially critical sociolinguistics, the role of statistics within a non-positivist linguistic theory, the existence of syntactic or pragmatic variables and the psycholinguistic significance of variable rules, are discussed in Sankoff (1988), which also contains the relevant bibliographical information. Most of the literature critical of variable rules is in fact directed towards the notion of a factor-conditioned linguistic choice process or these wider epistemological questions.

The development of variable rules themselves can be traced in Labov (1969, 715), Cedergren/Sankoff (1974, 333), Rousseau/Sankoff (1978c, 57), Sankoff/Labov (1979, 189), Sankoff (1985, 74) as well as the papers cited in connection with the specific topics covered in the preceding sections.

16. Literature (selected)

Bickerton, Derek (1973) "Quantitative versus dynamic paradigms: the case of Montreal *Que*", in: *New ways of analyzing variation in English*, Bailey, C.-J. N./Shuy, R. W., eds., Washington, D. C., 23–43.

Bishop, Yvonne M. M. (1969) "Full contingency tables, logits, and split contingency tables", in: *Biometrics* 25, 119–128.

Bishop, Yvonne M. M./Fienberg, Stephen E./Holland, Paul W. (1975) *Discrete multivariate theory*, Cambridge, Mass.

Cedergren, Henrietta J./Sankoff, David (1974) "Variable rules: performance as a statistical reflection of competence", in: *Language* 50, 333–355.

Cedergren, Henrietta J./Rousseau, Pascale/Sankoff, David (1986) "La variabilidad de /r/ implosiva en el español de Panamá y los modelos de ordenación de reglas", in: *Estudios sobre la fonología del español del Caribe*, Núñez Cedeño, R./Páez Urdaneta, I./Guitart, J., eds., Caracas, 13–20.

Hout, Roeland van (1984) "The need for a theory of choice in sociolinguistics", in: *Linguistische Berichte* 90, 39–57.

Kalbfleisch, James G. (1984) "Aspects of categorical data analysis", in: *Topics in applied statistics*, Chaubey, Y. P./Dwivedi, T. D., eds., Montreal, 139–150.

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

Mantel, Nathan (1979) "Multidimensional contingency table analysis", in: *American Statistician* 33, 93.

Rousseau, Pascale/Sankoff, David (1978a) "A solution to the problem of grouping speakers", in: *Linguistic variation: models and methods*, Sankoff, D., ed., New York, 97–117.

Rousseau, Pascale/Sankoff, David (1978b) "Singularities in the analysis of binomial data", in: *Biometrika* 65, 603–609.

Rousseau, Pascale/Sankoff, David (1978c) "Advances in variable rule methodology", in: *Linguistic variation: models and methods*, Sankoff, D., ed., New York, 57–69.

Sankoff, David (1985) "Statistics in linguistics", in: *Encyclopedia of the statistical sciences* 5, New York, 74–81.

Sankoff, David (1986) "Ordenamiento de reglas variables: /r/ implosiva en un dialecto puertorriqueño", in: *Estudios sobre la fonología del español del Caribe*, Núñez Cedeño, R./Páez Urdaneta, I./Guitart, J., eds. Caracas, 109–115.

Sankoff, David (1988) "Sociolinguistics and syntactic variation", in: *Linguistics: the Cambridge survey* IV, Newmeyer, F., ed., London, 140–161.

Sankoff, David/Labov, William (1979) "On the uses of variable rules", in: *Language in Society* 8, 189–222.

Sankoff, David/Rousseau, Pascale (1974) "A method for assessing variable rule and implicational scale analyses of linguistic variation", in: *Computers in the humanities*, Mitchell, J., ed., Edinburgh, 3–15.

Sankoff, David/Rousseau, Pascale (1980) "Categorical contexts and variable rules", in: *Papers from the Scandinavian symposium on syntactic variation*, Jacobson, S., ed., Stockholm, 7–22.

Sankoff, David/Rousseau, Pascale (1981) "Echelles et règles", in: *Variation Omnibus*, Sankoff, D./Cedergren, H. J., eds., Edmonton, 157–269.

David Sankoff, Montréal (Canada)

113. Varietätengrammatik

1. Sprachliche Variabilität und ihre Beschreibung

Natürliche Sprachen wie das Deutsche, Französische, Lateinische oder Chinesische sind nicht einheitlich, sondern sie setzen sich aus einer Vielzahl verwandter Varietäten – Dialekten, Soziolekten, Registern, Fachsprachen usw. – zusammen, die in manchen der sie kennzeichnenden Regelhaftigkeiten übereinstimmen, in andern hingegen nicht. Diese Variabilität stellt den Forscher, sofern er es nicht vorzieht, seine Bemühungen auf eine einzige – vielleicht besonders hoch angesehene – Varietät zu beschränken, vor eine dreifache Aufgabe:

(1) Er muß die einzelnen Varietäten in ihren verschiedenen sprachlichen Eigenschaften charakterisieren. Zu diesen Eigenschaften zählen nicht nur jene, die man gewöhnlich der Grammatik im engeren Sinne (Phonologie, Morphologie, Syntax) zurechnet, sondern auch lexikalische ebenso wie bestimmte Besonderheiten im kommunikativen Verhalten (z. B. die Wahl der Anredeform oder die Regeln des Rederechts).

(2) Er muß die einzelnen Varietäten in ihrem Verhältnis zueinander beschreiben, also inwieweit sie in bestimmten Eigenschaften übereinstimmen oder nicht übereinstimmen. Dabei ergeben sich zumindest drei Probleme. Zum ersten gibt es angesichts der eben erwähnten Bandbreite dieser Eigenschaften keine durchgängige Methode, um dies zu leisten. Zum zweiten hat die moderne Linguistik ihre Beschreibungsinstrumentarien größtenteils für homogene, idealisierte Sprachformen entwickelt; es bedarf daher zusätzlicher Mittel und Wege, um diese Instrumentarien für den Vergleich von Varietäten nutzbar zu machen. Und zum dritten sind die Unterschiede zwischen den einzelnen Varietäten oft kontinuierlich, d. h. zwei Varietäten unterscheiden sich oft nicht im Vorhandensein und Fehlen einer bestimmten Regel, sondern im Mehr oder Minder ihrer Anwendung.

(3) Schließlich muß er beschreiben, wie die innersprachliche Variabilität mit irgendwelchen außersprachlichen Faktoren zusammenhängt – Faktoren wie beispielsweise soziale Schicht, geographische Verteilung oder Besonderheiten der Redesituation, um nur drei der bekanntesten zu nennen (vgl. Nabrings 1981, Kapitel 3). Eine wichtige Dimension der Variabilität ist nicht zuletzt die Entwicklung in der Zeit – sei es eines Individuums (im Erst- oder Zweitspracherwerb) oder einer Sprachgemeinschaft. Welche Rolle diese und viele weitere denkbare außersprachliche Faktoren für die Variation tatsächlich spielen, läßt sich nicht a priori sagen, sondern es ist eine Frage der empirischen Analyse.