SSLA, **21**, 453–475. Printed in the United States of America.

DEPENDENCE AND INTERACTION IN FREQUENCY DATA ANALYSIS IN SLA RESEARCH

Hidetoshi Saito
Rikkyo University, Tokyo

In SLA research, dependence in frequency data is a prevalent problem (Hatch & Lazaraton, 1991). Researchers usually regard the data as being independent and subject them to statistical analyses. Another problem in frequency data analysis in SLA research is the exclusion of interaction terms. This is the case because of the use of chi-square analysis or the unpopularity of multifactorial frequency data analyses. This study investigates the violation of the "independent observation" assumption as well as the effect of including interaction terms in frequency data analysis. Reanalyzing a couple of published data sets, the paper argues in favor of using multifactorial frequency data analyses over multiple chi-squares in order to take into account dependence and interaction of the frequency data. A set of recommendations for SLA researchers will be provided when statistically analyzing frequency data in SLA research.

Dependence in frequency data has long been recognized (Guy, 1980; Hatch & Lazaraton, 1991) but never fully explored in the field of second language acquisition. The selection of the appropriate statistical procedure depends on whether the data are based on independence of observations. That is, in the case of frequency data, an assumption that must be taken into account is that "each observed frequency is generated by a different subject" (Gravetter & Wallnau, 1991, p. 374). In many SLA studies, however, this important assumption is violated such that the researcher collects multiple frequency counts

The author would like to thank Don Winford, Ron Clason, all the referees, and the editorial staff of *SSLA* for their insightful comments and encouragement. Also, great appreciation is extended to Sharon Ash for her help.

Address correspondence to: Hidetoshi Saito, Rikkyo University, Language Center, 3-34-1, Nishi-Ikebukuro, Toshima-Ku, Tokyo, 171, Japan; e-mail: saito11@rb3.so-net.ne.jp.

© 1999 Cambridge University Press 0272-2631/99 $9.50

https://doi.org/10.1017/S0272263199003046 Published online by Cambridge University Press

Table 1. Idea units recalled by passage type and prereading instruction

		Frequency of recall		
Passage type	Instruction	Recalled	Not recalled	Total
Problem solving	No frame	27	153	180
	Minimal frame	30	150	180
	Expanded frame	49	131	180
Descriptive	No frame	12	87	99
	Minimal frame	23	77	100
	Expanded frame	20	80	100
(Total)		161	678	839

Note. $N = 120$ (each of six conditions contains 20). Adapted from Lee and Riley (1990).

from one participant and lumps them together in one cell with those from other participants. When this assumption is violated in this way, the consequence of a statistical analysis of frequency data is a positively biased p value (Kennedy, 1992). In other words, the result is more likely to be statistically significant than in the case in which the assumption is not violated.

Another important issue in statistical analysis of frequency data pertains to the inclusion of interaction terms. For example, it seems true that extralinguistic factors, such as gender and occupation, are likely to interact with each other and with linguistic factors (Sankoff, 1988). Such interaction terms have not been fully investigated in SLA and sociolinguistic research. This paper explores these two analytical issues in frequency data analysis in SLA research: the effects of dependence and interaction.

FREQUENCY DATA ANALYSES IN SLA RESEARCH

The type of data discussed in this paper is *frequency data,* also called nominal data, categorical data, qualitative data, and tabular data. When used as a dependent variable (DV), frequency data are categorical. Examples are: whether students quit studying the target language or not (dichotomous DV), whether learners produce correct relative clauses or not (dichotomous DV), or which types of relative clause are produced (multinomial DV). In studies that involve these DVs, independent variables (IV) can also be categorical—for example, first-year or second-year students (dichotomous IV), male or female (dichotomous IV), treatment A, treatment B, or control (multinomial IV), and the like. When such data are tabulated, frequency or occurrence of the given category goes into each cell. Table 1 shows an example of a frequency (or contingency) table from a study by Lee and Riley (1990).

In their study of foreign language reading, Lee and Riley (1990) investigated the effects of prereading instruction on the degree to which foreign language learners were able to recall the content of a passage accurately. The data were obtained through a recall protocol procedure, in which the participants

https://doi.org/10.1017/S0272263199003046 Published online by Cambridge University Press

Table 2. Studies that reported dependent frequency data with statistical analyses in *Studies in Second Language Acquisition* (1978–1996)

	Statistical analysis		
Time of studies	Chi-square	VARBRUL/ Log-linear	Total
1978–1989	11	2	13
1990–1997	7	2	9

recalled the content of a passage immediately after they had read it. The two passages that were used differed in textual organization—one a collection of descriptions organization and the other a problem-solution organization. The participants received three different sets of prereading instructions: The no-framework group received no prereading instruction, the minimal-framework group was informed about the type of passage before reading it, and the expanded-framework group was given some information about the content as well as the type of passage they were about to read. Six groups, each consisting of 20 participants, were randomly assigned to one of the six conditions (2 Passage Types × 3 Prereading Instructions). A research question addressed in this study was whether the three groups differed in their recall of main ideas, or *top-level idea units*. In other words, the influences of two IVs (Passage and Instruction) on the number of idea units recalled (Times Recalled DV) were examined by way of frequency count.

When frequency data are involved in research, two types of statistical analysis are possible: chi-square and multifactorial frequency data analyses. Chi-square is more popular than the latter in SLA research (see Table 2). In the aforementioned case, one can perform multiple chi-square analyses (which is, in fact, what Lee & Riley, 1990, did). In Table 1, for instance, a couple of 3 (Instruction) × 2 (Times Recalled) chi-squares are possible for each Passage Type. Employing these analyses, however, is not appropriate in at least three respects.

First, the data are dependent. A particular participant may contribute to one cell more than others, causing within-cell dependence (explained below). In the upper left cell of Table 1, for example, a participant might have contributed 9 counts out of the total of 27 counts, whereas some others share none or very few out of 27 counts. This biased frequency distribution makes chi-square analysis more likely to generate statistical significance, resulting in a positively biased p value (Kennedy, 1992).

Second, there is a possibility of inflated risk of Type I error in the same way that Type I error is increased by running a number of t tests instead of running a factorial ANOVA (Lissitz & Chardos, 1975). Type 1 error is making an incorrect decision by which one rejects the null hypothesis when one should not. In particular, it is the case where one erroneously finds significant results. The likelihood of committing Type 1 errors is increased by employing

https://doi.org/10.1017/S0272263199003046 Published online by Cambridge University Press

Table 3. Relationship between research questions, variables, and statistical analyses

Research question	Dependent variable		Independent variable
	Categorical	Numeric	
Difference	chi-square; log-linear; logit model	*t* test; ANOVA	categorical
Prediction	logistic regression; VARBRUL	linear regression	categorical
	logistic regression; discriminant analysis	linear regression	numeric
	logistic regression	linear regression	mixed
Difference (repeated)	CATMOD	repeated ANOVA	categorical
Prediction (repeated)	event history analysis; survival analysis	time-series regression	mixed

Note. Only univariate (single DV) procedures are listed. Log-linear analysis does not make distinctions between IV and DV.

numerous *t* tests or chi-squares. An important reason why statisticians have developed factorial ANOVAs and multifactorial frequency data analyses is to avoid committing this error.

The third problem of running multiple chi-squares for Table 1 is that the interaction term cannot be tested. For example, a Passage Type by Instruction Interaction is possible in the data but cannot be tested by running chi-squares. Additionally, one may wish to see the effects of Passage Type statistically controlling for the effects of Instruction, which is not possible by chi-squares. Multifactorial frequency data analyses are a way to get around the second and third problems but not the first problem.

Various multifactorial frequency data analyses are seen in statistics: log-linear, logit model, logistic regression, event history analysis, and so forth. Table 3 summarizes the relationships between variable types and multifactorial frequency data analyses.

Log-linear and logit models are identical except that log-linear makes no distinction between IV and DV. The logit model is an ANOVA equivalent procedure for frequency data. Logistic regression deals with both categorical and continuous IVs and is a regression equivalent procedure for frequency data. The DV of logistic regression has to be categorical.

VARBRUL is one of the most often employed multifactorial frequency data analyses in SLA research.[1] VARBRUL is a special case of logistic regression in which all the IVs included are categorical. The case in which all the IVs are categorical is usually handled by the logit model, which has more to do with a research question of differences between groups or categories—an ANOVA-type question. On the other hand, the type of question that is addressed in studies using VARBRUL concerns finding a set of effective predictors or IVs—a regression-type research question.

Logistic regression is typically used for finding a set of factors that predicts

https://doi.org/10.1017/S0272263199003046 Published online by Cambridge University Press

a dichotomous DV. For example, one may want to predict who will be likely to get or not get lung cancer (dichotomous DV). The researcher might enter age, gender, years of smoking, and weights as predictive variables of getting lung cancer. The logistic regression procedure estimates the *b* coefficients of each variable (they are transformed into factor weights or probabilities in VARBRUL) that composes a predictive model of lung cancer. To find a set of good predictors, the aggregate models of IVs are evaluated in reference to some criteria (discussed below). To take an SLA research example, Adamson and Regan (1991) investigated Vietnamese and Cambodian immigrants' acquisition of the English suffix *-ing.* They entered speaking style, gender, grammatical category, and following phonological environment as a set of factors to predict the presence or absence of *-ing.*

A basic assumption of the VARBRUL, or logistic regression, analysis is that the probability of choosing one category over another, such as presence or absence of a linguistic token, can be generated from the multiplication of IV effects, rather than the simple addition of those effects. Since Cedergren and Sankoff's (1974) seminal paper, Sankoff has continuously improved the computer programs that are based on this analysis (Rand & Sankoff, 1990) and VARBRUL has become a very popular analytic choice in quantitative sociolinguistics. In SLA research, researchers have begun to employ the analytic tool and have shown its applicability to investigations of interlanguage variation (Adamson, 1988; Adamson & Regan, 1991; Bayley, 1994; Bayley & Preston, 1996; Preston, 1989; Young, 1988, 1991, 1993).

Although it is capable of testing interaction and avoiding an inflated alpha risk, VARBRUL has rarely been used for testing interaction terms primarily because of the program constraints.[2] What is more critical, however, is that neither chi-square nor VARBRUL can handle the problem of dependence in frequency data.

PROBLEMS OF DEPENDENCE AND INTERACTION IN SLA DATA

Problem of Dependence

An important assumption in subjecting frequency data to a statistical analysis is independent observation. Two types of independence should be taken into account when analyzing frequency data: within-cell and between-cell independence (Hatch & Lazaraton, 1991).

Within-cell independence is an assumption that each observation in a single cell has to be obtained from a different person (Gravetter & Wallnau, 1991; Hatch & Lazaraton, 1991). For example, the study in Table 1 used a total of 839 counts. In such a case, each of the 839 counts needs to be obtained from a different participant; that is, from 839 people. In this instance, however, they were obtained from 120 participants. The statistical analysis is biased because a particular individual is likely to contribute to a certain cell more than other

https://doi.org/10.1017/S0272263199003046 Published online by Cambridge University Press

participants. When this assumption is violated in such a way, the resultant *p* values are not exact; there may be a positive bias (Kennedy, 1992).

Herein lies a fundamental dilemma between the statistical assumption and research questions in SLA research. Multiple sampling of a linguistic token or response from speakers is often required to understand their current state of proficiency or process of SLA. It follows, then, that this violation of within-cell independence is prevalent in SLA research. Table 2 shows the tabulation of those studies in *Studies in Second Language Acquisition* from 1978 to 1996 that employed chi-square, VARBRUL, or log-linear and that violated the within-cell independence assumption.

According to Hatch and Lazaraton, in their description of this problem, "tradition appears to approve of this use of chi-square procedure" (1991, p. 407), although they do not recommend it. The implicit assumption behind this violated use is, of course, that each frequency count from a person is independent, even though there is no theoretical or empirical basis for assuming so. Justifying this implicit assumption is necessary as long as researchers continue using chi-square or VARBRUL with the lumping-together practice.

Between-cell independence is another assumption that must be taken into account when analyzing frequency data. For example, suppose that a researcher observes L2 learners producing relative clauses in specific situations. Suppose also that the same learners are exposed to the same situation three times over 3 months and the researcher counts the occurrence of relative clauses each time. In such a case, carry-over effects would be expected, because the learners are liable to remember the situations. This, in turn, affects responses in the following observation. In factorial ANOVAs, the within-subject or repeated-measure design deals with this problem. In multifactorial frequency analyses, procedures such as event history analysis, survival analysis, and CATMOD are known procedures to deal with the repeated measure design.[3] Although not explored in this paper, the violation of between-cell independence is certainly seen in SLA studies as well. The consequence of this violation again is to generate biased *p* values.

Problem of Interaction

Interaction effects are commonly found in factorial ANOVA analyses. An interaction can be present when the effects of one IV vary at the different levels of the other IV. For example, an association between the likelihood of volunteering in class (DV) and first language background (an IV) may not be constant at different levels of gender (another IV). A specific example of interaction in frequency data is given in the next section.

Sankoff (1988) discusses the possibility of interaction among variables in VARBRUL analysis. As he states, extralinguistic factors (e.g., age group and gender) may interact with other extralinguistic factors or linguistic factors. Indeed, such was the case for a study by Adamson and Regan (1991). In their study of Vietnamese and Cambodian immigrants acquiring *-ing,* they sus-

https://doi.org/10.1017/S0272263199003046 Published online by Cambridge University Press

pected that two of the IVs, Gender and Speaking Style, might interact. Because of a constraint of the VARBRUL program they used, the speaking style variable was arbitrarily eliminated from the analysis (see note 2). Thus, they did not test the interaction term.

There are other reasons why SLA researchers have rarely tested interaction terms in frequency data. First, the most commonly used frequency data analysis in SLA, chi-square analysis, is not capable of testing interactions or tables with more than two dimensions. Second, those multifactorial frequency analyses that are capable of testing interactions have not been widely used in the field. Of the studies tabulated in Table 2, only one study, which used log-linear analysis, tested interaction terms in the frequency data.

Testing interaction terms is important because, when a researcher is searching for a set of effective predictors of DV (regression-type question), adding interaction terms may help find a better set of predictors. It is also essential because, when a researcher is interested in a difference among groups (ANOVA-type question), testing an interaction term may result in a very different picture from testing only main effects. It is unfortunate that this importance has not been well understood in frequency data analysis in SLA research.

TESTING DEPENDENCE AND INTERACTION

In what follows, reanalyses of two published SLA studies illustrate the lurking effects of within-cell dependence and interaction. They also show the advantages of using two types of multifactorial frequency data analyses.

Beebe's (1983) Data

Beebe (1983) reported the effects of three IVs on accurate production of English by 19 Puerto Rican children. The IVs involved in the study appear in Table 4: Program (whether children were in a monolingual program or a bilingual program), Interviewers' Ethnicity (whether children were interviewed by a monolingual English speaker or a bilingual, Spanish–English, speaker), and Gender (male or female). The learners were interviewed once by a monolingual interviewer and once by a bilingual interviewer. Accuracy was measured by the number of correct or incorrect sentences calculated in T-units. Beebe was particularly interested in whether the two types of interviewers made any difference in the influences upon the learners' Accuracy, which was the DV.

No main effects or interaction were tested in the original study. Note that the following analysis only illustrates the test of main effects and interaction. Neither within-cell dependence nor between-cell dependence was controlled in the analysis, because sufficient information was not available for this example.

https://doi.org/10.1017/S0272263199003046 Published online by Cambridge University Press

Table 4. L2 learners' accuracy by program, interviewer, and gender

Program and Interviewer	Gender	Accuracy		Total
		Correct	Incorrect	
Monolingual				
Monolingual	Male	559	326	885
	Female	353	156	509
Bilingual	Male	269	77	346
	Female	269	62	331
Bilingual				
Monolingual	Male	288	169	457
	Female	411	310	721
Bilingual	Male	191	103	294
	Female	237	151	388
(Total)		2577	1354	3931

Note. Adapted from Beebe (1983).

Procedure. As explained earlier, a logit model is an ANOVA-equivalent frequency data analysis. It can deal with a multidimensional frequency data table and test main effects and interactions. In this example, three main effects (Program, Interviewer, and Gender) and four interactions (Program by Interviewer, Interviewer by Gender, Program by Gender, and Program by Interviewer by Gender) were tested with logit model analysis. The logit model analysis consists of two stages, residual log-likelihood ratio chi-square (residual L^2) analysis and component L^2 analysis.[4]

Residual L^2 analysis. As shown in Table 5, each variable is hierarchically included in logit models (a model here is a linear combination of variables). The best model is chosen by examining Akaike's information criteria (AIC) and *p* values associated with residual L^2. Each residual L^2 is calculated by taking the difference between the expected frequencies that each model generates and the observed frequencies in the original data. A goal of residual L^2 analysis is to find a best fitted model. That a model fits the data well means that the residual is small. In other words, the better the model is, the smaller the residual L^2 is, anticipating a nonsignificant associated *p* value. A smaller AIC value is also indicative of a better fit.

Component L^2 analysis. Component L^2 is calibrated by taking the difference between pairs of residual L^2s. A significant component L^2 indicates a large effect. The presence or absence of effects is determined by considering both residual and component L^2 analyses.

Results. Table 5 shows the results of residual L^2 analysis, and Table 6 shows those of component L^2 analysis. Residual L^2 analysis indicated that only three interactions achieved a good fit between the models and the observed data: Program by Gender, Interviewer by Gender, and Program by Interviewer by Gender, all of which were found to be nonsignificant. Of the three, how-

https://doi.org/10.1017/S0272263199003046 Published online by Cambridge University Press

Table 5. Logit-model (residual L^2) analysis of accuracy of sentences (D) by program (A), interviewer (B), and gender (C)

AIC	Source	Models	DF	ResidL^2	*p*
1060.602		Null	15	1058.602	.000
284.783		AB	10	272.784	.000
143.458		AB, AC	9	129.458	.000
138.786		AB, AC, BC	8	122.786	.000
102.251	Program	AB, AC, BC, AD	7	84.251	.000
72.556	Interviewer	AB, AC, BC, AD, BD	6	52.556	.000
74.350	Gender	AB, AC, BC, AD, BD, CD	5	52.350	.000
50.909		AB, AC, BC, AD, BD, CD, ABC	4	26.909	.000
38.069	Program by Interviewer	AB, AC, BC, AD, BD, CD, ABC, ABD	3	12.069	.007
28.247	Program by Gender	AB, AC, BC, AD, BD, CD, ABC, ABD, ACD	2	0.247	.879
30.230	Interviewer by Gender	AB, AC, BC, AD, BD, CD, ABC, ABD, ACD, BCD	1	0.230	.631
32.000	Program by Interviewer by Gender	AB, AC, BC, AD, BD, CD, ABC, ABD, ACD, BCD, ABCD	0	0.000	1.000

Note. All the models also contain A, B, C, D terms, which are abbreviated in this table.

Table 6. Logit-model (component L^2) analysis of accuracy of sentence (D) by program (A), interviewer (B), and gender (C)

Source	Models	DF	Comp L^2	*p*
Main effects				
Program	AB, AC, BC, AD	1	38.535	.000
Interviewer	AB, AC, BC, AD, BD	1	31.695	.000
Gender	AB, AC, BC, AD, BC, CD	1	0.206	1.000
Interactions				
	ABC	1	25.441	.000
Program by Interviewer	ABC, ABD	1	14.840	.000
Program by Gender	ABC, ABD, ACD	1	11.822	.000
Interviewer by Gender	ABC, ABD, ACD, BCD	1	0.017	1.000
Program by Interviewer by Gender	ABC, ABD, ACD, BCD, ABCD	1	0.230	1.000
(Total/Null)	AB, AC, BC	8	122.786	.000

Note. All the models also contain A, B, C, D terms, which are abbreviated in this table. Additionally, all the interaction models contain AB, AC, BC, AD, BC, and CD terms.

ever, component L^2 analysis supports only the presence of Program by Gender interaction, which was found to be significant. Further support of this finding is the smallest AIC value (28.247) in residual L^2 analysis. Thus, only the significant Program by Gender interaction effect was subjected to a follow-up analysis.

Following Rice (1992) and Kennedy (1992), log odds (the natural logarithm of the ratio of the number of correct T-units to that of incorrect ones in this

https://doi.org/10.1017/S0272263199003046 Published online by Cambridge University Press

case) of the interaction effect were plotted in Figure 1. Four pairs of multiple comparisons (called focused comparisons) were performed. The results indicated that the probability of producing accurate sentences by the learners in a monolingual program is constantly significantly higher than that of the learners in a bilingual program, regardless of gender and interviewers, except for one context. That is, when male students were interviewed by a monolingual interviewer, no significant difference was found, L^2 (1) = .069, $p > 1.0$. This is a clear indication of interaction. That is, the effects of one IV are not the same at the different levels of the other IVs.

Young's (1988, 1991) Data

Participants in Young's study were 12 Chinese learners of English as an L2. Each participant, a native speaker of either Cantonese or Mandarin Chinese, was interviewed for about an hour, first by native speakers of English and then by native speakers of Chinese (in English). The purpose of the study was to find a set of IVs that predicts the learners' production or nonproduction of the English plural *-s.* The 12 participants produced 1,564 linguistic tokens, of which 1,021 (65%) were marked for plural.[5] All of the tokens from 12 participants were lumped together and classified according to the following 10 IVs: (a) Interviewer (peers or native speakers of English),[6] (b) Proficiency (high-proficiency or low-proficiency group, based on TOEFL score), (c) Definiteness (definite or indefinite nouns), (d) Animacy (animate or inanimate nouns), (e) Position (nouns functioning as prenominal modifiers of an NP or nouns functioning as the head of an NP), (f) Syntax (nouns functioning as subjects, objects, complements, or adverbials in a sentence), (g) Preceding Phonological Segment (final segment of the stem being a vowel, sibilant, nonsibilant fricative, nasal, stop, or lateral), (h) Following Phonological Segment (vowel or glide, consonant or liquid, or pause), (i) Redundant Plural Marking within the NP (numerals, partitives, quantifiers, plural demonstratives, or no redundant marking), and (j) Noun-Verb Concordance (the NP containing *-s* being the subject of an inflected verb, an uninflected verb, a complement of *there is* or *there was,* or a complement of *there are* or *there were*). Using VARBRUL, Young tested the effects of these 10 IVs upon the principal DV, Presence or Absence of the English Plural Marker.

Procedure. To probe within-cell dependence in the data, an IV based on who produced the set of tokens (i.e., Participant variable) was entered. If this IV indicating who produced the set of tokens is found to account for the data well or to contribute to a better fitted model, it means that there is within-cell dependence in the data.

As for interaction, Young did not include any interaction terms in the original study. Because Young's study included extralinguistic factors (Proficiency and Interviewer), it is possible that they interact with other factors. For example, it is highly possible that Proficiency may interact with a linguistic factor.

https://doi.org/10.1017/S0272263199003046 Published online by Cambridge University Press

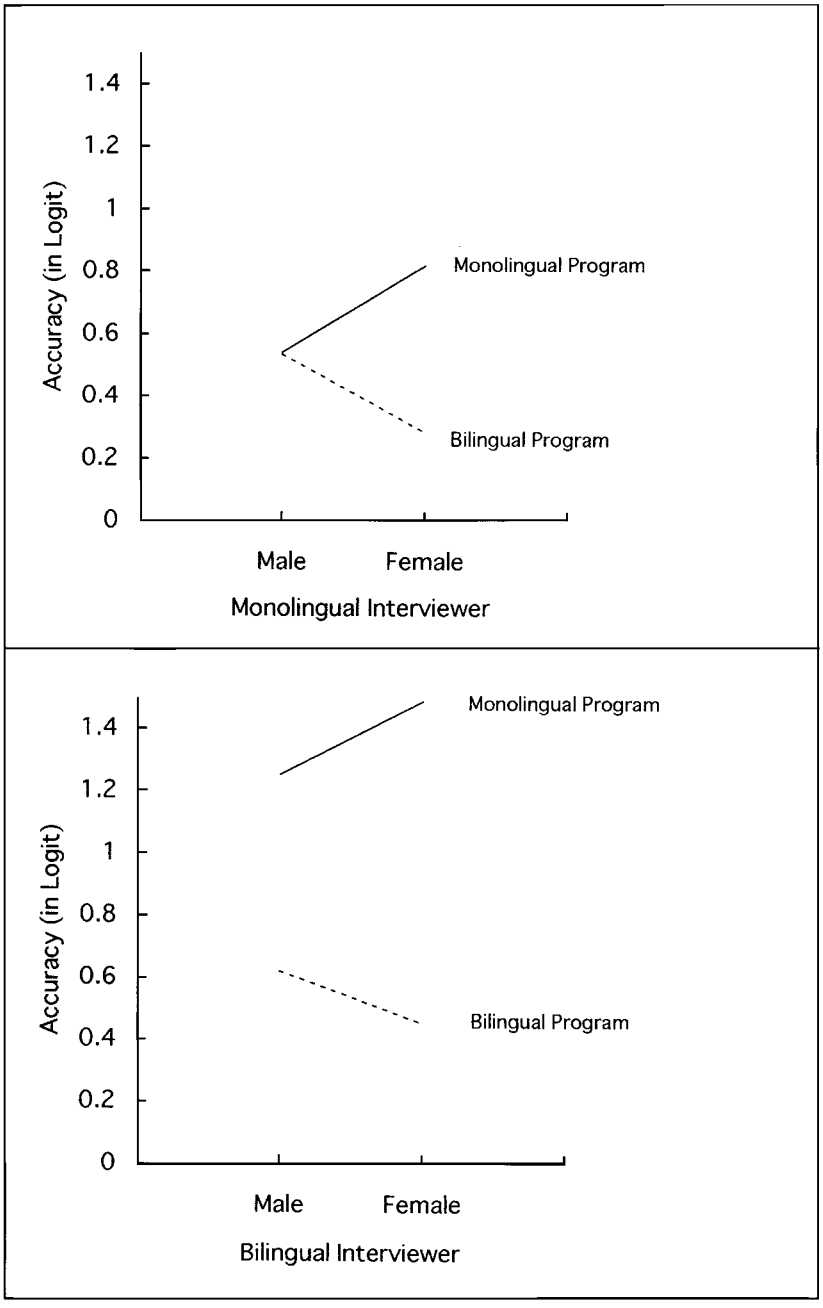


Figure 1. Interaction effect of Accuracy by Program and Gender stratified at each level of Interviewer.

https://doi.org/10.1017/S0272263199003046 Published online by Cambridge University Press

Table 7. Goodness-of-fit measures for logistic regression

Measure	Definition	Interpretation
Deviation L^2 (residual L^2)	A measure of the difference between the expected frequencies generated by a model and the observed frequencies. This is similar to a measure called total chi-square in VARBRUL.	If *nonsignificant,* the model fits the data well. If *significant,* the model fits the data badly.
Model L^2 (–2 log likelihood)	A test of the null hypothesis.	If *significant,* the model is better than the case where no IVs are included, which is the null model.
Percentages of correct prediction	Simple percentages of correct prediction or classification made by the model, indicating the extent to which the model is correctly able to predict the DV (Hosmer & Lemeshow, 1989).	The higher the correct prediction, the better the model.
Logistic Discriminating Function (LDF)	A point biserial correlation between expected frequencies generated by the model and observed frequencies in the data (Darlington, 1990).	The higher the coefficient, the better the model.
Lambda-*p* (Menard, 1994)	A proportional reduction in error measure, by which the amount of error in the prediction is reduced by the model.	The bigger the measure, the better the model.
R^2_L (Hosmer & Lemeshow, 1989)	Conceptually close to the adjusted R^2 in multiple regression—the portion of the variance explained by the model. It expresses a proportional reduction in the value of L^2. It is calculated by a ratio of the initial L^2 to the model L^2.	The bigger it is, the better the model.

In the following reanalyses, interactions between the Proficiency variable and linguistic variables were tested to see if the inclusion of an interaction term contributed to a better predictive model or accounted for more variance in the data.

Logistic regression analyses were run on the data.[7] Five different regression models (a model here meaning a linear combination of variables) were constructed. To compare the models, the criteria in Table 7, called goodness-of-fit measures, were used.[8] In addition to the six goodness-of-fit measures, the principle of Occam's razor was kept in mind. That is, all other things being equal, the fewer the factors in the model, the better the model (Guy, 1993).

https://doi.org/10.1017/S0272263199003046 Published online by Cambridge University Press

Table 8. Comparisons of five different logistic regression models of Young's (1988) data

	Models				
Feature	10 IVs	6-IVs-with-Proficiency	Participant	6-IVs-with-Participant	Interaction
IVs included	Proficiency* Interviewer Definiteness Animacy Noun position* Syntax* Preceding* Following** Redundancy* Concordance	Proficiency* Noun position* Syntax* Preceding* Following** Redundancy*	Participant*	Participant* Noun position* Syntax* Preceding* Following** Redundancy*	Proficiency Syntax* Preceding* Redundancy* Prof by Prec**
Deviation L^2	1778.426	1782.996	1879.667	1682.495	1782.508
DF	1540	1547	1553	1537	1545
p	.0000	.0000	.0000	.0053	.0000
Model L^2	243.400	238.830	142.159	339.330	239.318
DF	24	17	11	27	19
p	.0000	.0000	.0000	.0000	.0000
LDF (Rpbs)	.3832	.3801	.2980	.4548	.3821
Prediction %	70%	70%	67%	73%	70%
R^2_L	.1203	.1181	.0703	.1678	.1183
Lambda-p	.1378	.1470	.0422	.2261	.1470

Notes. Prof by Prec = Proficiency by Preceding interaction. See Appendix for B coefficients and standard errors of the variables. * = Wald statistics on B coefficients significant at $p < .00$; ** = $p < .05$

The goal of logistic regression analysis is to find a best fitted model in light of the measures. For a model to fit the data well, the expected frequencies generated from the chosen model must be very close to the observed frequencies of the original data. This is an indication that the model predicts the observed data well.

Results. Table 8 shows five different logistic regression models constructed from the Young data along with the six goodness-of-fit measures that were calculated for comparison. The first two columns display Young's original model (10 IVs), which includes all 10 IVs and a model (6-IVs-with-PROFiciency) composed of 6 IVs that were reported as significant in Young's study. The next column shows a model in which the Participant variable alone was included. The fourth model (6-IVs-with-PARTicipant) is identical to the 6-IVs-with-PROF model except that the Participant variable has replaced the Proficiency variable. The last model (Interaction) contains four IVs (Proficiency, Syntax, Redundancy, and Preceding) and one interaction term (Proficiency by

https://doi.org/10.1017/S0272263199003046 Published online by Cambridge University Press

Preceding). The last model was constructed by including and excluding several different interaction terms.

Let us closely examine each goodness-of-fit measure in Table 8. Deviation L^2 measures show that all the expected frequencies derived from the six models are significantly different from the observed data. This means that none of the models fit the data well. However, the 6-IVs-with-PART model is closest to nonsignificance ($p < .0053$), indicating the best fit among the five. For model L^2 measures, all the models are significant; all are better than the null model. LDFs reveal that the 6-IVs-with-PART model is the best of the five models; the expected frequencies of the Interaction model are correlated with the observed data at .45. Percentages of correct prediction are similar to each other. Nevertheless, the 6-IVs-with-PART model is the best again (73%). Note that the Participant variable alone can correctly predict 67% of the data. The R^2_L coefficient is a logistic regression analog of adjusted R^2. About 12% of variance is explained in both the 10 IVs model and the 6-IVs-with-PROF model, whereas 17% of variance is explained by the 6-IVs-with-PART model.[9] Furthermore, the Participant variable alone can account for about 7% of the variance. For lambda-*p* measures, the 6-IVs-with-PART model is best again; the highest error reduction, .23, is achieved by the model.

When comparing the Interaction model with the 6-IVs-with-PROF model, all the measures are almost indistinguishable. That is, the Interaction model is as well fitted as the 6-IVs-with-PROF model. On the other hand, the 6-IVs-with-PART model is significantly better in model L^2 measure than either that of the 6-IVs-with-PROF model or the Interaction model.

In summary, the 6-IVs-with-PART model is the best among the five models in terms of overall fit and predictive accuracy of the model. What is more, the Interaction model is no worse than the 6-IVs-with-PROF model in all measures.

DISCUSSION

The reanalysis of Young's data indicates that there was a within-cell dependence in the data. The Participant variable contributed to a fitted model (the 6-IVs-with-PART), which was better than the original 6-IVs-with-PROF model. The interaction between an extralinguistic factor (Proficiency) and a linguistic factor (Preceding) also contributed to a model as well fitted to the observed data as was the 6-IVs-with-PROF model. An interaction effect (Program by Interviewer) was also found in Beebe's (1983) data. In her data, however, within-cell dependence was not examined.

Considering the present results and the advantages of multifactorial frequency data analyses, the present researcher makes the following suggestions for analyzing frequency data.

1. When a frequency table has more than two dimensions (e.g., Tables 1 and 4), the best option for statistical analysis is multifactorial analysis. Multiple implementa-

https://doi.org/10.1017/S0272263199003046 Published online by Cambridge University Press

tions of chi-square analysis have disadvantages, as is the case for implementing a number of *t* tests: increasing alpha risks, being unable to examine a main effect while statistically controlling for other effects, and disregarding interaction effects.

2. When within-cell independence is violated, researchers should test the effect of the violation by including a participant variable, based on who produced a set of frequency counts. Including the participant variable helps, though not perfectly, to statistically control for the effect of within-cell dependence. Keep in mind that whenever this assumption is violated, the associated *p* value is not exact.
3. Researchers should test the interaction effects in frequency data. When interaction effects are considered in logistic regression, an analysis becomes cumbersome because the number of models that need to be tested explodes. This problem is similar to the "equivalent model" problem in the covariance structure modeling literature (MacCallum, Rosnowski, & Necowitz, 1992). For example, in the reanalysis of Young's data, it was found that the Interaction model is as good as the 6-IVs-with-PROF model. Which model should be chosen? Are there any other models that fit equally well? The only way to answer these questions is to obtain new data and retest the alternative models with the new data. What is more critical is that researchers recognize the existence of equivalent models and evaluate the models with the data at hand.
4. Each researcher should be aware of the statistical assumption that is violated. For example, random sampling is almost always violated in social science research. Thus, our generalizations and inferences are always limited. In addition to random sampling, what other assumptions are violated in a study? If within-independence is violated, that will further limit generalizations, inferences, and interpretation of statistical significance. Each researcher should discuss such limitations in a study.

Having said this, a few words of caution are in order. First, there are some concerns about simply leaving the domain of multiple chi-squares. There may be cases in which multiple chi-squares are more appropriate than multifactorial frequency data analyses. Research questions and research design take priority in the principal reasoning in choosing statistical procedures. Whatever choice is made, the researcher needs to justify why a certain procedure is chosen over others. If multiple chi-square is chosen, the researcher should explain why it is preferred, being aware of the disadvantages. Also, it is not the statistical procedure that solves all the problems, but research questions, carefully planned research design, and data collection procedures that take on more significant roles.

Second, interaction effects do not need to be investigated all the time. If a study bases itself on a theory that assumes no interaction whatsoever among the variables or if there are strong theoretical or empirical reasons to expect no interaction, it does not make sense to investigate interaction. In the field of SLA, many studies are exploratory with little theoretical basis. In such a case, it is advisable to explore the possibility of interactions among the variables. Again, the research question and theoretical basis associated with it take on more significant roles in deciding which main effects and interactions are included in statistical analyses.

https://doi.org/10.1017/S0272263199003046 Published online by Cambridge University Press

Finally, alternative approaches to chi-square or multifactorial analyses can be more appropriate in some cases. For example, in the Lee and Riley (1990) data in Table 1, the possible maximum number of idea units recalled in the problem-solving passage was nine. Thus, the data could have been considered as a nine-item test with dichotomous responses (recalled or not) by learners. Because the possible maximum number of idea units was five in the other passage, it can also be taken as a five-item test. The whole data set can now be considered as scores rather than frequencies and be subjected to a MANOVA with the scores of two different tests as DVs. In other cases, linguistic tokens produced by each participant can be transformed into percentiles. Because percentiles are ordinal data (Allen & Yen, 1979), they can be subjected to nonparametric procedures such as Kruskal-Wallis and Friedman. Furthermore, a researcher can set an a priori cutoff point for an IV in order to get around within-cell dependence. For example, an IV used in Pavesi's (1986) study was a dichotomous variable, the number of students who produced 80% correct relative clauses or those below 80%. The most critical reminder, however, is to avoid violations in designing a study; that is, before collecting data. It is equally important to question whether statistical analyses are necessary with all the violations at hand. Simple descriptive statistics, raw counts, and percentages of each participant may be sufficient in certain cases.

CONCLUSIONS

SLA research questions often require multiple sampling from a learner. By obtaining multiple samples of a learner's responses or linguistic tokens, researchers can pinpoint the current state of the learner's ability, proficiency, knowledge, or stages of SLA. Frequency data are often used as a critical indicator of such states and processes. Despite the obvious significance of multiple-sampled frequency data, there is no suitable statistical procedure for analyzing them as is. The consequence of the violated use is to generate inaccurate *p* values and to limit inferences and generalizations. On the other hand, disregarding interaction effects in frequency data can lead to missing better or equivalent models and differential patterns of effects that may lurk in the data.

What is argued in this paper is a modest proposal for SLA researchers. It is hoped that more researchers will be aware of these critical problems and make efforts to find ways to get around these problems. If the goal of any discipline is accumulation of knowledge, the knowledge accumulated must be correct. Inappropriate uses of statistical analysis and conclusions based on such uses hinder the correct interpretation of the data and result in accumulation of invalid knowledge, which further militates against the progress of the field.

(*Received 10 November 1997*)

https://doi.org/10.1017/S0272263199003046 Published online by Cambridge University Press

NOTES

1. VARBRUL has been developed in the field of quantitative sociolinguistics. Cedergren and Sankoff (1974) first introduced multifactorial frequency analyses—an application and a nonapplication model—based on Labov's (1969) additive model of the variable rule. The two models were further changed into a logistic regression model (Rousseau & Sankoff, 1978). This logistic regression model is what is now called VARBRUL, an acronym for both the variable rule and the statistical method.

2. Adamson and Regan stated that "the program assumes that no variables statistically interact" (1991, p. 13), which implies the program's incapability of testing interaction. Recent versions of VARBRUL seem to be able to handle interactions, although the procedure is not described in the first published guide of VARBRUL by Young and Bayley (1996). There are unpublished lecture notes on dealing with interaction using a VARBRUL program prepared by Beckman (1996).

3. Hatch and Lazaraton (1991) gave a brief presentation of the use of CATMOD. Incidentally, psychometric theories hold an assumption that one can avoid using the repeated measure frequency analyses. This assumption is called local independence, which means that a response to an item in a test does not affect a response to another item in the same test. This assumption is explicitly stated and statistically supported in item response theory.

4. Log-likelihood ratio chi-square is slightly different from Pearson's chi-square. Both are calculated from taking a difference between the expected frequencies that a model generates and the observed frequencies in the original data. But only L^2 involves natural logarithmic transformation. For detailed information of logit model analysis, readers are referred to Agresti (1996), Demaris (1992), and Kennedy (1992).

5. In the present reanalyses, 1,565 tokens were used based on the published data (Young, 1991). These data include one extra "absent" token in Jennifer.

6. In Young's 1988 paper, this factor had four categories. They were reduced to two categories in the subsequent book (Young, 1991).

7. An overview of the statistical estimation procedure and mathematical details of logistic regression is beyond the scope of the present paper. Interested readers are referred to introductory books such as Agresti (1996), Hosmer and Lemeshow (1989), Kleinbaum (1994), and Menard (1994).

8. When a linear (ordinary least square) regression model is employed, the created model is evaluated with two goodness-of-fit measures: *F* and *R*. The *F*-test generates statistical significance. *R* is a correlation between the model and the data. In logistic regression, two typical goodness-of-fit coefficients are also used: model L^2 and deviation L^2 measures, both of which are sample-size dependent. Statistical significance depends on sample size (see, for example, Keppel, 1991). The bigger the sample size, the more likely significance will be achieved. Thus, as pointed out by Menard (1994), a logistic regression model should also be assessed not only with model L^2 and deviation L^2 measures but with other indicators of model efficacy—something independent of sample size influence, just like R in linear regression. Researchers have proposed several sample-size independent goodness-of-fit measures for logistic regression such as LDF, lambda-*p*, and percentages of correct prediction.

9. This 5% increase in R^2_L is an improvement in model fit. Note that R^2_L is not an exact parallel to R^2 in linear regression. One may wonder if this 5% improvement is significant. The interpretation of this seemingly small improvement depends crucially on external factors: research question, research area, practical issues, previous studies, and so on. Even 3% improvement may make a difference in one context, although 10% may not be enough in another. This is similar to the fact that statistical significance does not always imply epistemological significance or practical significance. It depends on the contexts in which significance is interpreted.

REFERENCES

Note. *Study included in Table 2. †Participant factor considered in analyses.

Adamson, H. D. (1988). *Variation theory and second language acquisition.* Washington, DC: Georgetown University Press.

Adamson, H. D., & Regan, V. M. (1991). The acquisition of community speech norms by Asian immigrants learning English as a second language: A preliminary study. *Studies in Second Language Acquisition, 13,* 1–22.*

Agresti, A. (1996). *An introduction to categorical data analysis.* New York: Wiley & Sons.

https://doi.org/10.1017/S0272263199003046 Published online by Cambridge University Press

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole Publishing.
Bayley, R. (1994). Interlanguage variation and quantitative paradigm: Past tense marking in Chinese-English. In E. E. Tarone, S. M. Gass, & A. D. Cohen (Eds.), *Research methodology in second-language acquisition* (pp. 157–181). Hillsdale, NJ: Erlbaum.
Bayley, R., & Preston, D. R. (Eds.). (1996). *Second language acquisition and linguistic variation*. Amsterdam: Benjamins.
Beckman, M. (1996). *Notes on analysis of frequencies and logistic regression*. Unpublished lecture notes, Department of Linguistics, The Ohio State University, Columbus.
Beebe, L. M. (1983). Risk-taking and the language learner. In H. W. Seliger & M. H. Long (Eds.), *Classroom oriented research in second language acquisition* (pp. 39–66). Rowley, MA: Newbury House.
Blum-Kulka, S., & Levenston, E. A. (1987). Lexical-grammatical pragmatic indicators. *Studies in Second Language Acquisition, 9,* 155–170.*
Cedergren, H. J., & Sankoff, D. (1974). Variable rules: Performance as a statistical reflection of competence. *Language, 50,* 333–354.
Cohen, A. D. (1989). Attrition in the productive lexicon of two Portuguese third language speakers. *Studies in Second Language Acquisition, 11,* 135–149.*†
Darlington, R. B. (1990). *Regression and linear models.* New York: McGraw-Hill.
Demaris, A. (1992). *Logit modeling: Practical applications.* Newbury Park, CA: Sage.
Edge, B. A. (1991). The production of word-final voiced obstruents in English by L1 speakers of Japanese and Cantonese. *Studies in Second Language Acquisition, 13,* 377–393.*
Ellis, R. (1987). Interlanguage variability in narrative discourse: Style shifting in the use of the past tense. *Studies in Second Language Acquisition, 9,* 1–20.*
Gaise, S. J. (1982). Native speaker-nonnative speaker interaction among academic peers. *Studies in Second Language Acquisition, 5,* 74–81.*
Gravetter, F. J., & Wallnau, L. B. (1991). *Essentials of statistics for the behavioral sciences*. St. Paul, MN: West Publishing.
Guy, G. R. (1980). Variation in the group and the individual: The case of final stop deletion. In W. Labov (Ed.), *Locating language in time and space* (pp. 1–36). New York: Academic Press.
Guy, G. R. (1993). The quantitative analysis of linguistic variation. In D. Preston (Ed.), *American dialect research* (pp. 75–94). Amsterdam: Benjamins.
Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. Rowley, MA: Newbury House.
Hosmer, D. W., Jr., & Lemeshow, S. (1989). *Applied logistic regression.* New York: Wiley & Sons.
Huffines, M. L. (1991). Acquisition strategies in language death. *Studies in Second Language Acquisition, 13,* 43–55.*
Hulstijn, J. H., & Marchena, E. (1989). Avoidance: Grammatical or semantic causes? *Studies in Second Language Acquisition, 11,* 241–255.*
Ingram, E. (1978). Applied linguistics, linguistic research, and the empirical model. *Studies in Second Language Acquisition, 1,* 37–53.*
Kennedy, J. J. (1992). *Analyzing qualitative data: Log-linear analysis for behavioral research* (2nd ed.). New York: Praeger.
Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.
Kleinbaum, D. G. (1994). *Logistic regression: A self-learning text.* New York: Springer.
Labov, W. (1969). Contraction, deletion, and inherent variability of the English copula. *Language, 45,* 715–762.
Lee, J. F., & Riley, G. L. (1990). The effect of prereading, rhetorically-oriented frameworks on the recall of two structurally different expository texts. *Studies in Second Language Acquisition, 12,* 25–41.*
Lissitz, R., & Chardos, S. (1975). A study of the effect of the violation of the assumption of independent sampling upon the Type I error rate of the two-group *t* test. *Educational and Psychological Measurement, 35,* 353–359.
MacCallum, R. C., Rosnowski, M., & Necowitz, L. B. (1992). Model modification in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin, 111,* 490–504.
Major, R. C., & Faudree, M. C. (1996). Markedness universals and the acquisition of voicing contrasts by Korean speakers of English. *Studies in Second Language Acquisition, 18,* 69–90.*
Menard, S. (1994). *Applied logistic regression analysis.* Newbury Park, CA: Sage.
Olshtain, E. (1987). The acquisition of new word formation processes in second language acquisition. *Studies in Second Language Acquisition, 9,* 221–232.*

https://doi.org/10.1017/S0272263199003046 Published online by Cambridge University Press

Paradis, J., & Genesee, F. (1996). Syntactic acquisition in bilingual children: Autonomous or interdependent? *Studies in Second Language Acquisition, 18,* 1–25.*

Pavesi, M. (1986). Markedness, discoursal modes, and relative clause formation in a formal and an informal context. *Studies in Second Language Acquisition, 8,* 38–55.

Pica, T., & Doughty, C. (1985). The role of group work in classroom second language acquisition. *Studies in Second Language Acquisition, 7,* 233–248.*

Pica, T., Holliday, L., Lewis, N., Berducci, D., & Newman, J. (1991). Language learning through interaction: What role does gender play? *Studies in Second Language Acquisition, 13,* 343–376.*

Preston, D. (1989). *Sociolinguistics and second language acquisition.* Oxford: Blackwell.

Rand, D., & Sankoff, D. (1990). Goldvarb version 2.0 [Computer software]. Montreal, Canada: Centre de recherches mathématiques, Université de Montréal.

Rice, J. C. (1992). Loglinear analysis: Analysis of categorical variables in the logit setting. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 2, pp. 1–52). Greenwich, CT: JAI Press.

Robison, R. E. (1990). The primacy of aspect: Aspectual marking in English interlanguage. *Studies in Second Language Acquisition, 12,* 315–330.*

Rousseau, P., & Sankoff, D. (1978). Advances in variable rule methodology. In D. Sankoff (Ed.), *Linguistic variation: Models and methods* (pp. 57–69). New York: Academic Press.

Sankoff, D. (1988). Variable rules. In U. Ammon, N. Dittmar, & K. J. Mattheir (Eds.), *Sociolinguistics: An international handbook of the science of language and society* (pp. 984–997). Berlin: de Gruyter.

Schumann, J. H. (1987). The expression of temporality in basilang speech. *Studies in Second Language Acquisition, 9,* 21–41.*†

Tomlin, R. S. (1984). The treatment of foreground-background information in the on-line descriptive discourse of second language learners. *Studies in Second Language Acquisition, 6,* 115–142.*†

Tyler, A., & Bro, J. (1993). Discourse processing effort and perceptions of comprehensibility in nonnative discourse: The effect of ordering and interpretive cues revisited. *Studies in Second Language Acquisition, 15,* 507–522.*

White, L. (1987). Markedness and second language acquisition: The question of transfer. *Studies in Second Language Acquisition, 9,* 261–286.*

Wolfe-Quintero, K. (1992). Learnability and the acquisition of extraction in relative clauses and *wh*-questions. *Studies in Second Language Acquisition, 14,* 39–70.*

Young, R. (1988). Variation and the interlanguage hypothesis. *Studies in Second Language Acquisition, 10,* 281–302.*

Young, R. (1991). *Variation in interlanguage morphology*. New York: Peter Lang.

Young, R. (1993). Functional constraints on variation in interlanguage morphology. *Applied Linguistics, 14,* 76–97.

Young, R., & Bayley, R. (1996). VARBRUL analysis for second language research. In R. Bayley & D. R. Preston (Eds.), *Second language acquisition and linguistic variation* (pp. 253–306). Amsterdam: Benjamins.

https://doi.org/10.1017/S0272263199003046 Published online by Cambridge University Press

APPENDIX

Table A1. B coefficients and standard errors of each factor of five models in Table 8

	Models			
	10 IVs		6-IVs-with-PROF	
Independent Variables	B	*SE*	B	*SE*
Proficiency				
High	.5841	.0680	.5804	.0676
Low	–.5841		–.5804	
Interviewer				
Native	.0036	.0587		
Nonnative	–.0036			
Definiteness				
Definite	–.0900	.0788		
Indefinite	.0900			
Animacy				
Animate	.0503	.0765		
Inanimate	–.0503			
Noun position				
Prenominal	.9701	.3381	.9684	.3360
Head	–.9701		–.9684	
Syntax				
Subject	–.2550	.1929	–.2616	.1566
Object	–.3408	.1103	–.3062	.0996
Complement	.2650	.2134	.1805	.1740
Adverbial	.5229	.1266	.5398	.1191
Uncategorized	–.1922	.1618	–.1525	.1566
Preceding phonological environment				
Vowel	.1957	.1299	.1994	.1290
Sibilant	–.2991	.1863	–.3208	.1834
Nonsibilant fricative	.8048	.4082	.8510	.4050
Nasal	–.0938	.1567	–.1061	.1548
Stop	.2022	.1379	.2082	.1366
Lateral	–.8099	.2150	–.8318	.2129
Following phonological environment				
Vowel/glide	.2474	.0902	.2413	.0891
Consonant/liquid	–.2272	.0970	–.2303	.0962
Pause	–.0201	.0820	–.0110	.0816
Redundancy				
Numeral	.5629	.1895	.6109	.1832
Partitive	–.4857	.2190	–.4620	.2169
Quantifier	–.1810	.1778	–.1564	.1719
Demonstrative	.7335	.5429	.6276	.5317
No redundancy	–.6296	.1609	–.6201	.1594

https://doi.org/10.1017/S0272263199003046 Published online by Cambridge University Press

Concordance				
Subject of uninflected verb	.3693	.3363		
Subject of *-s* inflected verb	.2297	.4093		
Complement of *there are/were*	.1377	.4237		
Complement of *there is/was*	–1.1017	.7561		
Uncategorized	.3649	.2583		
Constant	1.2041	.4322	1.5355	.3694

Note. The so-called VARBRUL weights are calculable using

$$p = \frac{e^x}{1 + e^x},$$

where e is antilog and x is B coefficient.

https://doi.org/10.1017/S0272263199003046 Published online by Cambridge University Press

Table A2. B coefficients and standard errors of each factor of five models in Table 8

	Models			
	6-IVs-with-PART		Participant	
Independent Variables	B	*SE*	B	*SE*
Participant				
Ai	–.3805	.2834	–.0965	.2553
Chang	1.0215	.1730	.7089	.1576
Guo	–.6191	.1678	–.8163	.1491
Jennifer	–.2913	.2179	–.2616	.1930
Lily	–.8464	.2543	–.5348	.2163
Mary	–.6471	.1802	–.4324	.1580
Norman	–.4228	.2072	–.2951	.1884
Pearl	–.4774	.2222	–.4934	.2025
Quian	1.2879	.2100	.9394	.1966
Rainbow	.9562	.2147	.7066	.2008
Sally	1.1033	.1819	.8846	.1691
Wendy	–.6844	.2692	–.3095	.2374
Noun position				
Prenominal	.9342	.3397		
Head	–.9342			
Syntax				
Subject	–.3320	.1496		
Object	–.3120	.1045		
Complement	.1905	.1817		
Adverbial	.6281	.1250		
Uncategorized	–.1746	.1633		
Preceding phonological environment				
Vowel	.2012	.1347		
Sibilant	–.3353	.1921		
Nonsibilant fricative	.4879	.4179		
Nasal	.0623	.1633		
Stop	.2555	.1431		
Lateral	–.6717	.2227		
Redundancy				
Numeral	.6237	.1903		
Partitive	–.3699	.2276		
Quantifier	–.1193	.1773		
Demonstrative	.5812	.5391		
No redundancy	–.7157	.1633		
Following phonological environment				
Vowel/glide	.2800	.0932		
Consonant/liquid	–.1570	.1014		
Pause	–.1231	.0862		
Constant	1.6425	6.2723	.5608	.0591

https://doi.org/10.1017/S0272263199003046 Published online by Cambridge University Press

Table A3. B coefficients and standard errors of each factor of five models in Table 8

	Interaction	
Independent Variables	B	*SE*
Proficiency		
High	.3588	.7989
Low	–.3588	
Syntax		
Subject	–.3037	.1348
Object	–.3369	.0994
Complement	.1473	.1742
Adverbial	.4637	.1183
Uncategorized	.0295	.1502
Preceding phonological environment		
Vowel	–.0623	.8022
Sibilant	–.7286	.8183
Nonsibilant fricative	3.0057	3.9710
Nasal	–.3863	.8083
Stop	.0015	.8045
Lateral	–1.8300	.8620
Redundancy		
Numeral	.6587	.1845
Partitive	–.4733	.2162
Quantifier	–.1971	.1700
Demonstrative	.6478	.5238
No redundancy	–.6361	.1572
Proficiency by preceding interaction		
High group by vowel	.1796	.8017
Sibilant	.4546	.8180
Nonsibilant fricative	1.1446	3.9707
Nasal	.2296	.8081
Stop	.1413	.8045
Lateral	–.2415	.8622
Constant	.8954	.8100

Note. Interaction B coefficients of low group are the opposite signs of interaction B coefficients of high group.

https://doi.org/10.1017/S0272263199003046 Published online by Cambridge University Press