Pragmatics: Quantitative Methods

FRANCESCO ARCIDIACONO

Quantitative Methods and Pragmatics

Quantitative research has the goal to conceptualize the reality in terms of variables, to measure these variables, and to study the relationships between them (Punch, 2009). In the case of pragmatics (Ochs & Schieffelin, 1979; Green, 1989), the variables may be on categorical, ordinal, or interval scales, such as categorical variables that correspond to conversational or discursive strategies (Ochs & Schieffelin, 1983; Freedle, 1990). In pragmatics, different analyses can be applied to verify the inferential relations among variables, but in cases where there are three or more categorical variables multipath frequency analysis is especially useful.

The aim of this entry is to describe how to apply log-linear analysis as one type of multipath frequency analysis in order to explore pragmatic phenomena, using as an example data from a study on verbal conflicts in family conversations.

Overview of Verbal Conflicts and Log-Linear Analysis

Based on the belief that verbal conflicts among adults and children are educational tools of interaction and means of cultural socialization, this entry illustrates a quantitative method used to analyze these conversations (Arcidiacono & Pontecorvo, 2009). Parents–children's adversarial conversations (video-recorded during everyday interactions at dinnertime, and then transcribed turn by turn) were categorized in terms of the following variables which were seen as relevant dimensions of conflict (Labov & Fanshel, 1977; Vuchinich, 1990; Garvey & Shantz, 1992):

1. the orientation (with two dimensions: serious, non-serious),
2. the modality (with two dimensions: mitigated, aggravated), and
3. the closing (with three dimensions: withdrawal, submission, compromise).

Traditionally, the quantitative technique used to measure the relationships between two categorical variables was to compare observed frequencies to the frequencies expected by either chance or a hypothesis, and then to compute the difference with a Pearson chi-square statistic (χ^2) in which a large value indicated that there is a relationship between the variables. For more than two variables, several chi-squares could be computed to measure the relationships between variables 1 and 2, variables 1 and 3, and variables 2 and 3, for example. While perhaps providing insight into the main effects of one variable on another, no insights would be gained on possible interactions among the three variables. In order to test for the statistical significance of the main and interaction effects between variables, a multivariate extension to chi-square tests called hierarchical log-linear analysis is used in which models are applied to represent the logarithm of the expected cell frequency as a linear combination of the variables' effects.

The Encyclopedia of Applied Linguistics, Edited by Carol A. Chapelle.

DOI: 10.1002/9781405198431.wbeal0947

In our example, the following research question was addressed: What are the discourse patterns when parents activate the conflict with their children? We hypothesized an association between (a) serious and aggravated dimensions of the conflict with withdrawal as the closing strategy and (b) non-serious orientations and mitigated modalities with final compromises as the closing strategy.

An Example Study

Conducting the Analysis

A $2 \times 2 \times 3$ design was used in order to test three variables of verbal conflict and to explore the association among these factors and their levels. The data were analyzed using the statistical program BMDP Dynamic 7.0 (Brown, 1983). For this program, as the first step the researcher needed to type in syntax and to paste the control language (Table 1) at

Table 1 Control language

Syntax		*Description of codes and functions*
/ INPUT	TITLE IS “CONFLICT”.	Name of the file data
	VARIABLES = 4.	Number of variables (including frequencies)
	FORMAT IS FREE.	
/ VARIABLE	NAMES = O, M, C, FREQ.	O = orientation; M = modality; C = closing
/ CATEGORY	CODES (O) = 1, 2.	Codes or cutpoints are required if there are more than 10 values for any categorical variable
	NAMES (O) = S, N.	S = serious; N = non-serious
	CODES (M) = 1, 2.	
	NAMES (M) = MI, A.	MI = mitigated; A = aggravated
	CODES (C) = 1, 2, 3.	
	NAMES (C) = D, SO, CO.	D = withdrawal; SO = submission; CO = compromise
/ TABLE	INDICES = O, M, C.	Variables defining indices for tables
	COUNT = FREQ.	Variables containing frequencies when input is a cell indices and frequencies
	DELTA = 0.5	Value to be added to each cell before the analysis
/ FIT	MODEL = OMC.	Describes the log-linear models to be fit to data in a multiway table: the hierarchical model to be fitted
	ASSOCIATION = 3.	Estimates partial and marginal association for all effects of order
/ PRINT	LAMB.	Specifies types of tables to be printed: the table of estimated log-linear parameters and their standard errors
/ END		End of the input of control language

Table 2 Data set

Closing	*Modality*	*Orientation*		*Total*
		S	*N*	
D	MI	2	7	9
	A	20	5	25
	Total	22	12	34
SO	MI	0	0	0
	A	22	0	22
	Total	22	0	22
CO	MI	12	3	15
	A	11	0	11
	Total	23	3	26

a command prompt in order to code variables and dimensions. Using a "free" format, variables are separated by one or more blanks, a single comma, or both. The file is then saved as a command file (format.DAT) that in our case was named Conflict.DAT.

Then the researcher enters the data as text (or using a pre-existing file in the computer system). Table 2 reports the data set for our study (including frequencies of each variable under examination). At the end, this input file is saved as Conflict.INP. The program is thus ready to run and to analyze the data.

As mentioned before, the use of log-linear models aims at understanding the relationships between factors through a screening for an appropriate model, testing, comparing, understanding models under consideration, and examining cells with large disparities between observed and expected values under a chosen model. According to Marascuilo and Busk (1987) "multidimensional contingency tables can be examined in a unified way by using log-linear models in either a hypothesis testing mode or model-building approach" (p. 443). In the first case, the researcher sets up a number of null hypotheses to be rejected.

Table 3 Hierarchical models of interaction

Models	*Description*
[OMC]	Multivariate dependence: model of saturation
[OM][OC][MC]	Multivariate independence: dependence of three couples of variables
[OM][MC]	Marginal multivariate independence: dependence of two couple of variables
[OM][OC]	
[OC][MC]	
[OM][C]	Marginal multivariate independence: independence of one variable from the other two
[OC][M]	
[MC][O]	
[O]	H_0—Multivariate independence: same probability for each category
[M]	
[C]	

Table 4 Decisional models

K-FACTOR	*D.F.*	*LR CHISQ (G^2)*	*PROB*	*PEARSON CHISQ (χ^2)*	*PROB*	*ITER*
0-MEAN	11	92.15	0.00	92.14	0.00	0
1	7	43.33	0.00	41.71	0.00	2
2	2	0.43	0.80	0.47	0.79	11
3	0	0.00	1.00	0.00	1.00	1

In the second case, specific main effects and interactions are supposed to exist and interrelations among variables are tested for goodness of fit.

In our case, the strategy was to compare sets of models and then to define models based on hypothesis tests for individual parameters within a specific set. Table 3 shows seven possible hierarchical models of interaction computed in the study: from the more complex [OMC], named the dependent model of saturation, to the simple model [O][M][C] that postulates the independence between variables. In log-linear analysis "the idea is to find the least complex model that nonetheless generates expected values not too discrepant from the observed ones, as determined by a goodness-of-fit-test" (Bakeman & Gottman, 1986, p. 194); so, three steps in the model-building approach include (a) a backward analysis that begins with the saturated model, (b) a consideration of more parsimonious models, and (c) an examination of each bi-dimensional interaction and main effects.

Results of the Study

After running the statistical program with the control language and the input data, we obtain the file Conflict.OUT containing different records. First, we consider the decisional table models (Table 4).

At this stage, the likelihood ratio chi-square G^2 (Marascuilo & Busk, 1987; Rojewski & Bakeman, 1997a, 1997b) is used as the associated statistical test in order to explore the decisional models and to test for goodness of fit. Starting from each specified model, the program deletes in turns each simple or multiple effect. Through backwards elimination the probability associated to the difference between the saturated model and the model that fits better is considered. By this procedure, the program automatically screens all possible models in a generating class hierarchy for the most parsimonious one. Starting with the saturation model is not required: whatever model the researcher starts with, the backward elimination algorithm drops the least useful term one step at a time, stopping when the deleted effect is significant. In a design with K variables, the K-factor interaction parameter ensures that the hierarchical structure of the model is preserved. The goodness of fit of a log-linear model is tested using the likelihood ratio statistic G^2 (instead of the Pearson chi-square statistic χ^2) because G^2 can be partitioned into unique components that have the additive property similar to the sum of squares of the analysis of variance. In other words, G^2 is additive under positioning for nested models and its p value has to be > 0.05 for a well-fitting model. At this stage, the expected frequencies need to be tabulated. For a three variables model, we use an iterative proportional fitting algorithm. This procedure permits adjustments to fit each of the marginal sub-tables specified in the model. In Table 4, the algorithm indicates the highest significance level for the K-factor = 2. Having dropped the three-way term, the remaining generating models are two-way interactions and main effects.

We now consider Table 5, in which marginal and partial association are simultaneously used to screen various interactions, and to determine whether they are necessary, not necessary, or questionable. Different G^2 statistics are assessed for each combination of the

Table 5 Association options

Effect	*Partial association*				*Marginal association*			
	D.F.	*CHISQ(G^2)*	*Prob*	*Iter*	*D.F.*	*CHISQ(G^2)*	*Prob*	*Iter*
O	1	32.83	0.00					
M	1	13.48	0.00					
C	2	2.51	0.29					
OM	1	12.69	0.00	2	1	10.94	0.00	2
OC	2	12.54	0.00	2	2	10.80	0.00	2
MC	2	21.16	0.00	2	2	19.41	0.00	2
OMC	2	0.43	0.80					

Table 6 Ratio of the log-linear parameter estimate to its standard error

			Orientation	
	Modality		S	N
Closing		MI	–2.66	2.66
	MI	A	2.66	–2.66
D	–0.53	0.53	–2.29	2.29
SO	–1.50	1.50	0.57	–0.57
CO	2.37	–2.37	1.18	–1.18

variables. As both tests lead to the same association options about the effects to be considered, no problem exists in choosing partial or marginal association (if either the partial or marginal test is not significant, judgments can be made according to the researcher's knowledge about the variables).

In our case, significant G^2 values concern the two-way interaction models.

Once estimates of expected frequencies are obtained, the effect parameter estimates are produced and related to odd and odd ratios. The models of two-way interactions [OM][OC][MC] are then examined (Table 6). The estimates of the effects are similar to the calculation of main and interaction effects in a factorial analysis of variance. In our case we only consider the interaction effects because they override main effects. Odds are described as the ratio between the frequency of being in one category and the frequency of not being in that category. In order to investigate the quality of fit of the mode further, we evaluate the individual cell residuals that show why a model fits or display a lack of fit. This process involves standardized residuals for each cell. The cells with the largest residuals show where the model is least appropriate.

Generally, odd ratios above 1 indicate positive association among variables, while odd ratios smaller than 1 indicate negative association. If the value is 1, the variables are not associated. In our case, significant associations concern a confidence interval of z-scores ≥ 1.96 and $p < 0.05$. In particular, the interaction between orientation and modality (i.e., the four cells at the top right side in Table 6) reveals an association between the serious orientation and the aggravated modality, while the mitigated modality is associated with the non-serious orientation ($z = 2.66$); the interaction between orientation and closing (i.e., the six cells at the bottom right side in Table 6) confirms the association between the non-serious orientation and the withdrawal (recall that withdrawal was coded as "D" in the dataset) in closing conflicts ($z = 2.29$); the interaction between modality and closing (i.e., the six cells on the left side in Table 6) shows a significant association between the compromise and the mitigated modality to close a conflict ($z = 2.37$).

Results indicated that one discourse pattern used when parents activate verbal conflicts with their children was explained by a model associating the serious orientation to the aggravated modality (i.e., first hypothesis), but not necessarily followed by a withdrawal. When parents engage in a non-serious way, conflicts develop through a mitigated modality, in turn associated to a compromise as closing (i.e., second hypothesis). Another pattern indicates the interaction between the non-serious orientation and the withdrawal as closing.

Conclusion

As this analysis highlighted some discursive patterns generated by specific models, other possible lines of investigation (Arcidiacono, in press) have been devoted to specifically explore the cases in which children activate verbal conflicts with parents.

The use of log-linear models is useful to investigate association structures between a set of categorical variables, because "one of the advantages in using the log-linear models and the associated G^2 statistics is that there are many ways that one can partition the likelihood ratio statistic" (Marascuilo & Busk, 1987, p. 454). In applying the log-linear analysis, the use of the statistical program BMDP Dynamic 7.0 is based on a simple control language and a clear output. However, there are alternative options, such as ILOG, R, SAS, and SPSS software programs that offer practical solutions to combine different variables' effects available to be interpreted by the researcher. In the case of quantitative methods in pragmatics, we also encourage an awareness of other different models for analyzing contingency tables that are based on the binomial or multinomial nature of a response variable (see also Goodman, 1964; Marascuilo, 1966; Marascuilo & McSweeney, 1977).

SEE ALSO: Child Pragmatic Development; Comparing Groups With Multiple Independent Variables; Multiple Regression

References

Arcidiacono, F. (in press). "But who said that you eat when you want and what you want?" Verbal conflicts at dinnertime and strategic moves among family members. In J. P. Flanagan & A. M. Munos (Eds.), *Family conflicts: Psychological, social and medical implications*. New York, NY: Nova Science.

Arcidiacono, F., & Pontecorvo, C. (2009). Cultural practice in Italian family conversations: Verbal conflict between parents and preadolescents. *European Journal of Psychology of Education, 34*(1), 97–117.

Bakeman, R., & Gottman, J. M. (1986). *Observing interaction: An introduction to sequential analysis*. New York, NY: Cambridge University Press.

Brown, M. B. (1983). Two-way and multiway frequency tables. Measures of association and the log-linear model (complete and incomplete tables). In W. J. Dixon (Ed.), *BMDP Statistical Software* (pp. 143–206). Berkeley: University of California Press.

Freedle, R. (1990). *Advances in discourse processes: Conversational organization and its development*. Norwood, NJ: Ablex.

Garvey, C., & Shantz, C. U. (1992). Conflict talk: Approaches to adversative discourse. In C. U. Shantz & W. W. Hartup (Eds.), *Conflict in child and adolescent development* (pp. 93–121). Cambridge, England: Cambridge University Press.

Goodman, L. A. (1964). Simple methods for analyzing three-factor interaction in contingency tables. *Journal of the American Statistical Association, 59*, 319–52.

Green, G. M. (1989). *Pragmatics and natural language understanding*. Hillsdale, NJ: Erlbaum.

Labov, V., & Fanshel, D. (1977). *Therapeutic discourse: Psychotherapy as conversation*. New York, NY: Academic Press.

Marascuilo, L. A. (1966). Large sample multiple comparisons. *Psychological Bulletin, 65*, 280–90.

Marascuilo, L. A., & Busk, P. L. (1987). Log-linear models: A way to study main effects and interactions for multidimensional contingency tables with categorical data. *Journal of Counseling Psychology, 34*(4), 443–55.

Marascuilo, L. A., & McSweeney, M. (1977). *Nonparametric and distribution-free methods for the social sciences*. Monterey, CA: Brooks.

Ochs, E., & Schieffelin, B. B. (1979). *Developmental pragmatics*. New York, NY: Academic Press.

Ochs, E., & Schieffelin, B. B. (1983). *Acquiring conversational competence*. London, England: Routledge.

Punch, K. F. (2009). *Research methods in education*. Thousand Oaks, CA: Sage.

Rojewski, J. W., & Bakeman, R. (1997a). Applying log-linear models to the study of career development and transition of individuals with disabilities. *Exceptionality, 7*(3), 169–86.

Rojewski, J. W., & Bakeman, R. (1997b). Log-linear modeling and analysis: Reflections on the use of multivariate categorical data in social science research. *Exceptionality, 7*(3), 199–203.

Vuchinich, S. (1990). The sequential organization of closing in verbal family conflict. In A. D. Grimshaw (Ed.), *Conflict talk* (pp. 118–39). Cambridge, England: Cambridge University Press.

Suggested Readings

Bakeman, R., & Robinson, B. F. (1994). *Understanding log-linear analysis with ILOG. An interactive approach*. Hillsdale, NJ: Erlbaum.

Gnisci, A., & Bonaiuto, M. (2003). Grilling politicians. Politicians' answers to questions in television interviews and courtroom examinations. *Journal of Language and Social Psychology, 22*(4), 385–413.

Knoke, D., & Burke, P. J. (1980). *Log-linear models*. Newbury Park, CA: Sage.

Maroni, B., Gnisci, A., & Pontecorvo, C. (2008). Turn-taking in classroom interactions. Overlapping, interruptions and pauses in primary school. *European Journal of Psychology of Education, 23*(1), 59–76.

Melamed, I. D. (2000). Models of translational equivalence among words. *Computational Linguistics, 26*(2), 221–49.

Merialdo, B. (1994). Tagging English text with a probabilistic model. *Computational Linguistics, 20*(2), 155–71.