Loglinear Models for the Analysis of Language Acquisition Data

Ping Li

University of Richmond
pli@richmond.edu

1. Introduction

While it is common practice for researchers in psychology and other social sciences to use inferential statistical methods such as t-test, F-test, and chi-square test, it is only the beginning for linguists and investigators of language acquisition to get acquainted with the full power of inferential statistics. As the linguistic science becomes more quantitative, there have been admirable efforts to introduce statistics into the field, with the publications of several statistics books specifically designed for linguists (Anshen, 1978; Butler, 1985; Hatch & Farhady, 1982; Woods, Fletcher, & Hughes, 1986; see Grotjahn, 1988, for a review). However, none of these books has discussed a very important statistical method for the analysis of categorical data, the loglinear analysis — a method that has nevertheless been applied widely in sociology and other social sciences. In this paper, I will first examine a problem that many researchers in language acquisition may have encountered, i.e., the limited power of the analysis of categorical data by the use of chi-square. I will then discuss how loglinear analysis overcomes the problem. Although descriptions about loglinear analysis are available in many statistics books, they are in general not easily accessible to language acquisition researchers because of their technicality and mathematical flavor. For this reason, I will deliberately avoid very technical descriptions here, but will instead present the rationale behind the method, the basic procedures involved in the analysis, and a real example that makes use of this method to illustrate the significance of loglinear analysis for language acquisition data..

2. Limitations of Chi-Square

Chi-square (χ^2) test, like the analysis of variance (ANOVA), is one of the most frequently used statistic methods in child language studies. One reason for the popularity of this method is that most child language data

Journal of Cognitive Science 3 : 27-41, 2002.

involve observations classified into different categories, and the observations are represented by frequency counts of a given behavior under a given condition for a given group of children, and thus, chi-square test, a goodness-of-fit test, is the most suitable method. It tests how well under given conditions the observed frequencies (F_O) fits the expected frequencies (F_e) on the basis of the null hypothesis that there are no differences between groups or conditions. If the discrepancy between F_O and F_e is large enough with respect to the relevant degrees of freedom (*df*), then we can reject the null hypothesis and entertain our research hypothesis that there are differences. But if the discrepancy is small, then we cannot reasonably reject the null hypothesis.

Table 1 The Use of *-ed* vs *-ing* in English-Speaking Children's Early Productive Speech: Hypothetical Data (1)

Verb Types	Inflectional Markers		Total
	-ed	*-ing*	
Resultative	43	62	105
Non-resultative	17	38	55
Total	60	100	160

In most cases when we use chi-square we are concerned with the relationship between the individual classification variables, that is, with whether they are independent of each other or whether they interact to affect the outcome of our observations. A simple example here can illustrate this. Suppose that we want to investigate English-speaking children's acquisition of verbal inflections. We are interested in whether there is a difference between the use of the past tense marker *-ed* and the use of the progressive marker *-ing* in children's early productive speech, and whether this difference is associated with the semantic categories of verbs (for empirical studies of this kind, see Bloom, Lifter, Hafitz, 1980; Harner, 1981; McShane & Whittaker, 1988; see Li and Shirai 2000, for a review). Suppose further that we collected data from a total of 160 children, in which each child had one utterance describing a given enacted situation. The verbs in children's utterances could be classified into two types, resultative versus non-resultative, according to whether the verb indicates an end result (e.g., *break*) or not (e.g., *walk*). Counting the occurrences of *-ed* and *-ing* in these utterances, we could get the results in Table 1.

If the null hypothesis for this set of data is true, then we would expect that F_e in a particular cell (i.e., the combination of a particular morpheme with a particular verb) to be the product of its marginal totals divided by the grand total (for example, the cell occupied by 43 should be $(105 \times 60)/160 = 39.38$). If all the observed frequencies F_O are the same as or similar to their corresponding expected frequencies, then there is evidence that the variables are fully independent of each other in the outcome of the observations. A chi-square test performed on the data in Table 1 indicates that we cannot reject the null hypothesis: the use of *-ed* versus *-ing* in children's early productive speech is independent of the types of verb with which they are used ($\chi^2 = 1.55$, $df = 1$, $p > .05$). There is no statistical evidence that children's acquisition of the past tense marker and the progressive marker is associated with the semantic properties of verbs.

Table 2 The Use of *-ed* vs *-ing* in English-Speaking Children's Early Productive Speech: Hypothetical Data (2)

Verb Types	3-YEAR-OLDS		5-YEAR-OLDS		Total
	-ed	*-ing*	*-ed*	*-ing*	
Resultative	58	47	43	62	210
Non-resultative	12	43	17	38	110
Total	70	90	60	100	320

However, the application of chi-square test is greatly limited if we have a design that is just slightly more complex than the example described above. In the example given above, we were concerned with what is called a "2 × 2 contingency table", in which there were only two classification variables involved (2 inflectional markers by 2 verb types). If we were to study more than two classification variables, we would have a problem with this simple chi-square method. For example, in addition to the use of *-ed* versus *-ing* with respect to verb categories, we are interested in the developmental patterns for the acquisition of these morphemes, and so we need to have a third variable "age" in our design. To take a simple case, let us test only two age groups: 3-year-olds versus 5-year-olds, under the same conditions as in Table 1. Thus, we could have a new 2 × 2 × 2 design (2 inflectional markers by 2 verb types by 2 age groups), as in Table 2.

In principle, one could run two separate chi-square tests on the different age groups to test whether the use of *-ed* versus *-ing* is independent of the verb types in children's speech. And if there are more age groups, one can run more separate individual tests. However, doing so risks a higher "Type I Error" (rejecting the null hypothesis when it is actually true). This is because if we carry out two separate chi-square tests at the .05 significance level, the probability that at least one test will lead to Type I Error has increased to $1 - (.95)^2 = .097$, rather than .05. The more separate tests we do, the more likely it is for us to make a Type I Error.

For the analysis of data with more than two variables, chi-square is insufficient also because it cannot locate the exact patterns of effects (interaction effects as well as main effects) from different variables. Chi-square only tests the hypothesis of the independence between variables, while most often researchers want to know about both the main effects contributed by individual variables and the interaction effects contributed jointly by two or more variables, as is often shown by analysis of variance (it is obvious that the kind of data above cannot be tested using analysis of variance, since they are categorical in nature; see Howell, 1999). In view of these limitations of chi-square, we see loglinear analysis as a great tool to study individual effects and joint effects of classification variables with categorical data.

3. Basic Principles of Loglinear Analysis

Loglinear analysis has in recent years become a very important technique for the analysis of complex cross-classified categorical data. A number of references are available to researchers in the social and behavioral sciences (Bishop, Fienberg & Holland, 1975; Christensen, 1997; Everitt, 1977; Gilbert, 1981; Green, 1988; Kennedy, 1992; Knoke & Burke, 1980; Marascuilo & Busk, 1987). In loglinear analysis, a major task is to search for models that fit the observed data. The key procedure is to specify and compare loglinear models. A model, in this context, is a hypothesis or a conceptual framework about the observations in the data.

Each model represents the observed frequencies as realizations of some underlying probabilities. For example, the following independence model involves two variables and it specifies that the probability of a given frequency at the *ij*-th cell (represented as F_{ij}) is the product of the marginal probabilities ($p_{i.}$ and $p_{.j}$).

$$F_{ij} = p_{i.} \; p_{.j} \qquad (1)$$

The probability in this model can also be expressed as the sum of the marginal probabilities by the use of natural logarithms on both sides of the equation, as in the following:

$$\log F_{ij} = \log p_{i.} + \log p_{.j} \qquad (2)$$

Model (2) is thus a loglinear model in which the logarithms of the probabilities form a linear combination. Much like in multiple regression or analysis of variance, loglinear analysis breaks down frequency probabilities into additive components. In this way, the contributions of each variable to the data and its interaction with other variables can be evaluated precisely by the comparison between one model and another. It is in this sense that loglinear analysis is a model-testing method.

Before going on to look at how we do loglinear analysis, let us examine the general relationship among the models when compared in loglinear analysis. Loglinear models are hierarchically organized, differing in the number and complexity of the effects they include: some include only main effects, others main effects plus association effects (relationship between two variables), and still others main effects plus association effects plus interaction effects (relationship between three or more variables).[1] In the hierarchical loglinear models introduced here, models with higher-order effects presuppose the inclusion of the corresponding lower-order effects; that is, if a model includes an association effect, it must also include the corresponding main effects, and if it includes an

[1] In loglinear terminology, a distinction is made between association (two-way interaction) and interaction (three-way or higher-order interactions). See Gilbert (1981) for details.

interaction effect, it must also include the corresponding association effects and main effects.

To illustrate the difference between alternative loglinear models, consider a two-dimensional table where the classification variables are named A and B. There are five possible loglinear models for data in such a table.

$$\log F_{ij} = u \quad (3)$$
$$\log F_{ij} = u + u_{i.} \quad (4)$$
$$\log F_{ij} = u + u_{.j} \quad (5)$$
$$\log F_{ij} = u + u_{i.} + u_{.j} \quad (6)$$
$$\log F_{ij} = u + u_{i.} + u_{.j} + u_{ij} \quad (7)$$

Model (3) is a model with no variable effect: the probability of a frequency is the same for every cell (u represents the overall mean effect). Models (4) to (6) are models of main effect, in which $u_{i.}$ represents the main effect of variable A at the i-th cell, and $u_{.j}$ the main effect of variable B at the j-th cell. Model (7) is the association model, which incorporates the overall mean effect, the main effects of both A and B, and the association effect (represented as u_{ij}) of A and B (see Everitt, 1977 and Gilbert, 1981, for detailed procedures for the calculation of u-terms). Because model (7) includes all possible effects of a 2 × 2 design, it is also known as the *saturated model* for a data set with two classification variables.

The above general principles of loglinear analysis are the same for data with more than two variables. Although it is generally the case that the more variables there is the harder it becomes to explain the effects, loglinear analysis provides a convenient way to locate the detailed patterns of main effects, associations, and interactions by partitioning them in the above manner. Unlike in chi-square test, in loglinear analysis we do not need to carry out individual tests separately. It allows us to test all the effects in one shot. In the following discussion, main effect models will be designated as {A} for a variable named A, {B} for variable {B}, and so on. Models of association or interaction will be designated as {AB} where variable names are A and B, and so on. The model of no effect is not of direct interest in loglinear analysis, and thus will not be designated with a

label. The four models in (4) to (7) correspond to {A}, {B}, {A}{B}, and {AB}, respectively, in our new notation.

Like chi-square, loglinear analysis is also a goodness-of-fit test. It attempts to determine which models best fit a given set of data. The algorithm used in loglinear analysis also generates expected cell frequencies for each model, and, accordingly, it produces goodness-of-fit statistics (commonly the likelihood ratio statistic, designated as L^2 below) that indicate how well each model fits the data. Different models are then compared to determine which effects are responsible for the differences between the expected and observed frequencies. To evaluate whether a given effect contributes significantly to the data, the investigator has to compare models that include the effect with models that do not. The comparison is done by taking two models at a time and subtracting the L^2 value of one from the L^2 of the other, and the degrees of freedom (*df*) of one from the *df* of the other, and using the critical values of the chi-square distribution to evaluate the significance of the residual L^2 relative to the residual *df* (see example below).

In evaluating the various loglinear models for their fit to the data, we generally use two criteria: adequacy and parsimony. Conventionally, models with a significance level of .05 or above are considered to provide an adequate fit to the data.[2] All adequate models are acceptable as providing adequate accounts for the data. An alternative criterion for determining a good fitting model is that a model is good when the L^2 value is about equal to its degrees of freedom (Green, 1988). However, only the one that accounts for the most of the effects in the data and at the same time is also parsimonious will be considered the best model. There are two basic procedures to approach the best model: backward elimination and forward selection. In backward elimination, the researcher starts from the

[2] Special attention should be paid here to the interpretation of p values of L^2. In loglinear analysis, a p value above .05 indicates that a given model fits the data adequately. Thus the smaller the L^2 is relative to the *df* and the closer the p value is to 1.00, the better the model fits the data. This interpretation differs from the customary one for a p value, say, of χ^2 according to which values of *less* than .05 are taken to indicate a significant effect. See Knoke and Burke (1980:30-31) for discussion.

most complex model, i.e., the saturated model, and eliminates effects from it one by one, in a stepwise fashion (much like in multiple regression). Thus, for the models listed in (3) to (7), we first look at model (7), and go on with model (6) to see if it provides an adequate fit to the data when we eliminate the association effect u_{ij}. If model (7) is acceptable, we can then go on further to model (5) and (4), in which one of the main effect is eliminated. Model (3) will not be considered, since we generally cannot expect a model with no effect to account for our data. The other procedure, forward selection, is the reverse of backward elimination. In this way, we start with the simplest model, (4) or (5), in which only one main effect is incorporated, and then go on to see if other more complex models fit the data better. For a reasonably large set of data, the compromise between adequacy and parsimony will usually lead us to a model that is not too simple, but also not unnecessarily too complex.

4. An Illustration of Loglinear Analysis with Acquisition Data

In this section, I use a real example to illustrate how we can apply loglinear analysis to language acquisition data. The data come from a study on young Chinese-speaking children's acquisition of aspect markers (Li, 1990; Li and Bowerman, 1998). The question concerned was whether the acquisition of different aspect markers in Chinese is associated with children's knowledge of different semantic categories of verbs, and what the developmental pattern is for the acquisition process. Similar questions concerning the acquisition of tense and aspect in other languages have provoked much debate in the literature (e.g., Bloom et al, 1981; Bloom & Harner, 1989; Rispoli, M., & Bloom, L., 1985; Smith & Weist, 1987; Weist, R., Wysocka, H., Witkowska-Stadnik, K., Buczowska, E., & Konieczna, E., 1984; see Li and Shirai for a summary). Li examined the problem with several studies, including an elicited production experiment, which is discussed here. In the production experiment, 3- to 6-year-old kindergarteners were asked to describe some situations acted out with toys by the experimenter. The enacted situations fell into different event types in order to elicit verbs with different semantic properties in children's speech, for example: a duck was swimming, a doll was planting a tree, and a monkey was standing on a table (see Li, 1990; Li and Bowerman, 1998 for more details of the experiment). The results of this experiment are

shown in Table 3, which presents the frequencies of occurrence of the three aspect markers used with each verb type, broken down by age group. The observed frequency in each cell represents the number of combinations for a given aspect marker with a given verb type from all the utterances produced by children in that age group.

Table 3 The Use of Aspect Markers in Chinese-Speaking Children's Early Productive Speech

	3-YEAR-OLDS				4-YEAR-OLDS			
Verb Types	*-ne*	*zai*	*-le*	Total	*-ne*	*zai*	*-le*	Total
Process	32	13	21	66	45	14	19	78
Resultative	0	0	71	71	0	1	85	86
Telic	2	0	20	22	1	0	18	19
Punctual	11	4	5	20	15	7	7	29
Stative	16	3	15	34	23	2	17	42
	5-YEAR-OLDS				6-YEAR-OLDS			
	-ne	*zai*	*-le*	Total	*-ne*	*zai*	*-le*	Total
Process	30	24	18	72	52	26	7	85
Resultative	0	1	71	72	0	1	103	104
Telic	0	1	20	21	0	2	20	22
Punctual	16	15	9	40	20	13	3	36
Stative	24	11	14	49	29	4	6	39

* *-Ne* and *zai* are imperfective markers in Chinese, equivalent to English *–ing* in usage, while *-le* is the perfective marker, equivalent to English -*ed* or have+-*ed* (see Li, 1990 for more detailed discussion).

There are immediate problems to conduct chi-square tests on such a set of data. First, for this set of data, we are mostly interested in the interaction between the variables in addition to the roles of each individual variable. Chi-square does not allow us to achieve this goal. Second, even if we are interested in the independence relationship between the aspect markers and the verb types, we cannot legitimately run four separate chi-squares, one for each age group: we are likely to commit Type I Error with a probability as high as .19 (i.e., $1 - (.95)^4$) even if we are to carry out four separate chi-square tests. Finally, the data are inappropriate for chi-square test because there are many empty cells with zeros. This empty-cell situation is not uncommon in language data, where by definition there may be cases that should not have responses (such as the combination between the progressive marker zai and resultative verb, which is prohibited in the Chinese grammar).

Table 4 Loglinear Analysis On a Child Language Dataset (from Table 3)

Mode	Effect name*	*df*	L^2	*P*
(1)	{M}	57	1010.00	.00
(2)	{A}	56	1256.92	.00
(3)	{V}	55	1024.46	.00
(4)	{M} {A}	54	999.20	.00
(5)	{V} {M}	53	766.75	.00
(6)	{V} {A}	52	1013.66	.00
(7)	{V} {M} {A}	50	755.95	.00
(8)	{MA}	48	975.78	.00
(9)	{VM}	45	76.78	.00
(10)	{VA}	40	1002.55	.00
(11)	{MA} {V}	44	732.52	.00
(12)	{VM} {A}	42	65.98	.01
(13)	{VA} {M}	38	744.83	.00
(14)	{VM} {MA}	36	42.56	.21
(15)	{VA} {MA}	32	721.41	.00
(16)	{VM} {VA}	30	54.87	.00
(17)	{VM} {MA}{VA}	24	15.21	.91
(18)	{VMA}	0	0.00	1.00

* M = Aspect marker
V = Verb type
A = Age

Loglinear analysis solves these problems: in addition to its relaxation on empty cells,[3] it is able to handle multiple variables and is thus not subject to Type I Error in the above manner. Most important, it is ideal for locating exact patterns of effects of different variables, whether it be the main effect or the interaction effects. Table 4 presents the results from loglinear analysis performed on the data in Table 3. The results in the table were calculated with the SPSS-X HILOGLINEAR program (SPSS Inc, 1988; 1994).

Looking down the column for the *p* values in Table 4 we can see that only models (14), (17), and (18) provide an adequate fit to the data because only these models are above the significance level of .05 (see earlier discussion). To find out the best model, let us start from the most complex model and examine backward (i.e., using the backward elimination procedure). Model (18) is the saturated model, which, by definition, fits the data perfectly and so is uninteresting. Model (17), which incorporates all possible association effects among the three variables, fits nearly perfectly ($p = .91$), but this model does not give us a clear picture of which effects are most important since all association effects and main effects are included in the model. Model (14), which omits the association effect of verb type by age, also provides an adequate fit ($p = .21$), showing that the interaction between verb type and age does not account for much of the data structure. Among the three models that provide adequate fit, model (14) is thus selected as the best model since it provides an adequate fit and at the same time it is also parsimonious in terms of its employment of effect terms. Model (14) indicates that there are clear interactions between verb types and aspect markers and between aspect markers and age, and that these interaction effects (plus the main effects of verb type, aspect marker, and age) account sufficiently for the variations in the data.

To determine further which effects are important for the data, we need to examine not only the models that do provide an adequate fit, but also those that do not. Note that the key procedure in applying loglinear analysis is to compare the L^2 values of paired models and evaluate the

[3] Loglinear analysis sets empty cells to structural zeros, that is, with expected value of zero, and accordingly subtracts one degree of freedom for each of the cells.

significance of the difference between the two values. The models to be compared in a pair are models that are identical except one effect: one model includes a given effect while the other does not. Let us adopt the forward selection procedure here to determine which effects are important. For example, model (4) includes only the main effects of {M} and {A}, model (8) includes the corresponding association effect of {M} by {A}. To see if model (8) fits the data better than model (4), we subtract the L^2 and the *df* of model (8) from the corresponding values of model (4) and evaluate the significance of the difference with the chi-square distribution. As can be seen, the reduction of L^2 from model (4) to (8) is significant, given the cost of 6 degrees of freedom (χ^2 = (999.20- 975.78) = 23.42, *df* = (54-48) = 6, $p < .001$). Therefore, we conclude that the interaction between age and aspect markers accounts significantly better than the two main effects alone. When we compare in the same manner model (5), which incorporates only the main effect of {V} and {M}, and model (9), which also includes the association effect of {V} by {M}, we find an even more significant improvement of fit from (5) to (9) (χ^2 (8) = 690, $p < .001$). Our analyses indicate that overall, the use of aspect markers by Chinese-speaking children is differentially associated with different verb types. Finally, model (10), which includes the association effect of {V} by {A}, does not represent a significant improvement in fit over model (6) which includes only the main effects of {V} and {A} (χ^2 (12) = 11.1, $p > .05$). This suggests that in contrast to the other two association effects, the association of verb type by age does not account for much of the data variation.

Further comparisons between the simple association models (8) to (10) and the association plus main effect models (11) to (13) indicate that in all cases the inclusion of the remaining main effects is significant. However, model (12) represents only no significant improvement over model (9) in which the main effect of age is not included (χ^2 (3) = 10.8, $p > .05$), as compared with models (11) and (13). This shows that given the interaction between verb types and aspect markers, the differences between different age groups in the use of aspect markers are not all that great. Put it in a different way, it means that the association effect of {V} by {M} is by far the most important effect in the data. As can be seen in the above analysis, the association between verb type and aspect marker is more important than any other two-way association (e.g., verb type by age or aspect marker by age). Models that do not include the {V} by {M} relationship simply cannot capture the structure in the data; for example, although model (15) {VA} {MA} (L^2 (32) = 721.41, p = .00) incorporates two other association effects, it does not fit the data any better than the simpler model (9) (L^2 (45) = 76.78, p =.00).

Our discussion in this paper presents only a brief overview of the basic principles and the use of loglinear analysis for language acquisition studies. Although the example we examined above involves only three variables, loglinear analysis can be easily applied to categorical data involving four or more variables (see the reference list). Loglinear analysis is now incorporated in a great number of statistical packages, such as SPSS, BMDP, and SAS, many of which are commercially available. There are also programs specifically designed for loglinear analysis, such as ECTA (Everyman's Contingency Analyzer), GLIM (Generalized Linear Modeling), and LOGLIN (cf. Gilbert, 1981). Once the user has some basic knowledge of computerized data analysis and loglinear models, these specific programs can be easily followed.

Acknowledgement

The writing of this paper was made possible by a Faculty Research Grant from the University of Richmond, and in part by a visiting fellowship from the University of Hong Kong. I would like to thank Dirk Vorberg who convinced me of the power of loglinear models and guided me through data analyses during my dissertation research supported by the Max Planck Institute for Psycholinguistics, Nijmegen. I would also like to thank the anonymous reviewer for providing many helpful comments on an earlier version of the paper.

References

Anshen, F. (1978). *Statistics for linguists.* MA.: Newbury House Publishers, Inc.

Bishop, Y., Fienberg, S., & Holland, P. (1975). Discrete multivariate analysis: theory and practice. Cambridge, Mass.: MIT Press.

Bloom, L. & Harner, L. 1989. On the developmental contour of child language: A reply to Smith & Weist. *Journal of Child Language*, 16, 207-216.

Bloom, L., Lifter, K., & Hafitz, J. (1980). Semantics of verbs and the development of verb inflection in child language. *Language*, 56, 386-412.

Butler, C. (1985). *Statistics in linguistics*. Oxford: Blackwell.

Christensen, R. (1997). *Log-linear models and logistic regression*. Berlin: Springer-Verlag.

Everitt, B. (1977). *The analysis of contingency tables* . Chapman and Hall.

Gilbert, G. (1981). *Modelling society: An introduction to loglinear analysis for social researchers*. London: George Allen & Unwin.

Green, J. (1988). Loglinear analysis of cross-classified ordinal data: Applications in developmental research. *Child Development*, 59, 1-25.

Harner, L. (1981). Children talk about the time and aspect of actions. *Child Development*, 52, 498-506.

Hatch, E., & Farhady, H. (1982). Research design and statistics for applied linguistics. Rowley, MA: Newbury House.

Howell, D.C. (1999). *Fundamental statistics for the behavioral sciences* (4th ed.). Duxbury Press..

Grotjahn, R. (1988). Introducing (applied) linguists to statistics: A review of two books and some general remarks. *Studies in Second Language Acquisition*, 10, 63-68.

Keppel, G. (1982). *Design and analysis: A researcher's handbook*. Prentice-Hall, Inc.

Kennedy, J. J. (1992). Analyzing qualitative data: Log-linear analysis for behavioral research. 2nd Edition, N.Y.: Praeger.

Knoke, D., & Burke, P. (1980). *Log-linear models* . Sage Publications.

Li, P. (1990). *Aspect and aktionsart in child Mandarin*. Ph.D. dissertation, Leiden University.

Li, P. & Bowerman, M. (1998). The acquisition of grammatical and lexical aspect in Chinese. *First Language*, *18*, 311-350.

Li, P., & Shirai, Y. (2000). *The acquisition of lexical and grammatical aspect.* Berlin & New York: Mouton de Gruyter.

Marascuilo, L., & Busk, P. (1987). Loglinear models: A way to study main effects and interactions for multidimensional contingency tables with categorical data. *Journal of Counseling Psychology*, 34, 443-455.

McShane, J., & Whittaker, S. (1988). The encoding of tense and aspect by three- to five-year-old children. *Journal of Experimental Child Psychology* , 45, 52-70.

Rietveld, T., & van Hout, R. (1990). Loglinear analysis and weighted least squares. Manuscript, University of Utrecht.

Rispoli, M., & Bloom, L. (1985). Incomplete and continuing: Theoretical

issues in the acquisition of tense and aspect. *Journal of Child Language*, 12, 471-474.

Smith, C., & Weist, R. (1987). On the temporal contour of child language: A reply to Rispoli & Bloom. *Journal of Child Language*, 14, 387-392.

SPSS Inc. (1988). *SPSS-X user's guide.* Third Edition. New York: McGraw-Hill.

SPSS, Inc. (1994). *SPSS 6.1 base system user's guide*. SPSS, Inc. IL: Chicago

Weist, R., Wysocka, H., Witkowska-Stadnik, K., Buczowska, E., & Konieczna, E. (1984). The defective tense hypothesis: On the emergence of tense and aspect in child Polish. *Journal of Child Language*, 11, 347-374.

Woods, A., Fletcher, P., & Hughes, A. (1986). *Statistics in language studies.* Cambridge: Cambridge University Press.