USING MULTILEVEL MODELING IN THE ANALYSIS OF EXPERIMENTAL DATA: CUMULATIVE EFFECTS OF STRUCTURAL PRIMING IN CHILDREN

Filip SMOLÍK

Institute of Psychology, Academy of Sciences of the Czech Republic
Politických vězňů 7, 110 00, Praha 1, Czech Republic
E-mail: smolik@praha.psu.cas.cz

Abstract: Multilevel modeling is a flexible alternative to the traditional factorial ANOVA approach in the analysis of experimental data with repeated measures. This article describes a psycholinguistic experiment and provides a detailed account of the data analysis, demonstrating the use of multilevel models to include a continuous predictor and complex assumptions about error variance. The experiment investigated the effects of structural priming on reaction times in a word monitoring task. Pairs of sentences with identical or different syntactic structures were presented to 4- and 5-year-old children, whose task was to respond to a word presented in the second sentence. Multilevel modeling analysis revealed an interaction between the experimental condition and position of the trial within the experiment: the reaction times in the same-structure condition decreased over the course of the experiment, while they increased in the different-structure condition. The analysis demonstrates how multilevel models can be used to detect change in responses over the course of an experimental session.

Key words: syntactic priming, word monitoring, multilevel models, reaction time

Psycholinguistic experiments that use reaction times as the dependent variable usually compare mean reaction times in different experimental conditions using repeated-measures ANOVA. This method, based on the least squares estimation, has a number of limiting assumptions. One particular limitation concerns the extension of ANOVA that includes continuous predictors, i.e. ANCOVA. This method can only provide meaningful results if the effects of the continuous covariate are the same across the categorical conditions. This is one of the reasons why the use of continuous independent variables is infrequent in the analysis of experimental data and is often replaced by converting the continuous variable to a categorical factor with levels such as high vs. low score on the continuous variable. This type of conversion, however, leads to a loss of information and is strongly discouraged on methodological grounds (Cohen, 1983; MacCallum et al., 2002). An alternative method of data analysis is the use of mixed models, also known as hierarchical linear models or random coefficient models (Bryk, Raudenbush, 1992; Pinheiro, Bates,

Acknowledgments: The data analyzed in this article were collected while the author was affiliated to the Language Acquisition Studies Laboratory at the University of Kansas in Lawrence. The analysis and writing of this article was supported by the Internal start-up project No. Z70250504-I057 "Syntactic priming and language acquisition" awarded to the author by the Academy of Sciences of the Czech Republic.

This article was accepted before May 1, 2008, in accordance with previous publishing rules.

2000). This article presents a detailed account of a multilevel analysis of an experimental study with reaction time as the dependent variable. Besides the inclusion of a continuous independent variable in the analysis, the article demonstrates different approaches to error variance and methods of dealing with skewed data and outliers.

Multilevel Modeling in the Analysis of Experimental Data

Continuous independent variables may be useful in a number of circumstances. Many experiments use stimuli that differ in characteristics other than the experimental manipulation of interest; for instance, a word has a certain frequency, number of phonemes, number of phonological neighbors, or number of semantically related words. All these properties can have an effect on the reaction to an item, such as reaction time. In order to eliminate these variables as confounds, sets of stimuli in different experimental conditions must use stimuli with similar properties. This may be difficult, for instance because the number of potential stimuli is often limited. Inclusion of the covariate in the analysis provides a means of statistical control for potential confounds. Also, the effects of a continuous independent variable may be of interest as such. A number of studies exist that examine the relationships between word processing variables and word frequency. In this case, methods for dealing with continuous predictors are needed.

One continuous variable is of special interest since it is present in all psycholinguistic experiments that consist of multiple trials (items). The individual trials in an experiment are presented in a temporal sequence and may be influenced by practice or fatigue effects. Moreover, it is possible that the practice or fatigue effects have different impact in different experimental conditions, i.e. that the experimental condition interacts with the trial position. A straightforward method of accounting for these potential effects is to include the sequence number of a trial as a continuous covariate in the analysis. Due to the limitations of ANCOVA, the use of mixed models is a method of choice for this task.

Multilevel modeling extends the ANOVA/ANCOVA model so that it easily reflects the grouping present in the data, such as groups of observations coming from the same individual. ANOVA estimates the effects of independent variables on the dependent variable, and the remaining variance is considered residual. In mixed models, the group-specific (random) effects are estimated as well, and the estimation of independent variable effects (fixed effects) takes into account the differences between groups of observations. Residuals in mixed models are the difference between the actual measurement and the combined effect of group and independent variable. An example may be a participant whose responses in one experimental condition are 50 ms faster than in the other condition, but all responses from this participant are about 200 ms slower than the response time of other participants. The mixed model estimates separately the general deviation of the participant from the average reaction time (200 ms in the example), and thus reduces the error variance for estimating the systematic effect of condition (50 ms).

Besides their good ability to incorporate continuous predictors, multilevel methods can explicitly model some assumptions about the error variance. In standard ANOVA or regression techniques, as well as in most multilevel models, it is assumed that a single error variance term describes the residuals from all data points, regardless of the grouping. However, it is often the case that people, especially children, show large differences in the variability of their responses. Multilevel modeling makes it possible to assume different residuals in different groups of observations, e.g. in different participants. This reflects the intuitive fact that different people show different amounts of variation in their responses. It also improves the chance to identify outliers, such as extremely long reaction times, in the data. These are often diagnosed and removed using a fixed criterion that is applied to all observations (Ratcliff, 1993; Van Selst, Jolicoeur, 1994, cf.). However, this approach does not take into account individual differences in the overall value of observations and their variance. What is an extremely slow response for one person, thus a likely outlier, may be within the range of typical variance in another person. A multilevel model accounts for the variability in general speed by means of random effects, and it can account for the differences in intra-individual variability by means of subject-specific residuals.

Structural Priming in Children

The study presented in this paper investigated the phenomenon of structural priming in children. The term structural, or syntactic, priming refers to the fact that a recently processed syntactic structure has an influence on subsequent language production or comprehension. If such an influence can be measured, it provides evidence that the syntactic structure is internally represented at some level of the internal language system. This is of particular relevance in child language research: the phenomenon can be used to determine the level of abstractness in linguistic representations at various stages of language development.

Bock (1986) was the first to demonstrate the phenomenon experimentally in English-speaking participants, though similar phenomena had been reported earlier on the basis of analyses of spontaneous language use (Levelt, Kelter, 1982; Estival, 1985). Participants in Bock's study were asked to repeat sentences and provide verbal picture descriptions within one experiment. The study found that people were more likely to produce a passive description of a picture if they had repeated a passive prime sentence immediately before, compared to a neutral prime. A similar effect was found for actives, as well as for prepositional and double-object datives. The phenomenon has been replicated numerous times, including in languages other than English (Hartsuiker, Kolk, 1998), and in experiments using different response collection methods (Pickering, Branigan, 1998; Potter, Lombardi, 1998; Fox Tree, Meijer, 1999). A few studies explored and found facilitation effects of syntactic priming on reaction times (Smith, Wheeldon, 2001; Scheepers, 2003; Corley, Scheepers, 2002; Noppeney, Price, 2004).

There are two alternative theories of the mechanism responsible for structural priming. The authors arguing for the activation account (Pickering, Branigan, 1998; Pickering et al., 2000) assume that the phenomenon is due to the activation of structural nodes that represent sentential constituents. The opposing explanation is based on implicit learning (Chang et al., 2000, 2006). The structural priming in this view reflects the operation of a continuous learning process that builds the syntactic structures and leads to their generalization. The two approaches make different predictions about the time dynamics of structural priming. According to the activation account, syntactic priming should be a short-lived effect and should not last over intervening trials (Branigan et al., 1999, 2000). The implicit learning account is more consistent with persistent effects of priming, and with the cumulative effects that increase with increasing the number of presented syntactic structures of a particular type (Kaschak et al., 2006). The analyses presented in this article were performed in order to test for the possible presence of cumulative syntactic priming effects.

METHOD

The data analyzed in this article come from an experimental study of children's sensitivity to grammatical phenomena. The present paper analyzes the results of one component of the experiment in the younger group of participating children.

The experiment used the word-monitoring task to elicit timed responses. The task required the participants to respond by pressing a button when they heard a particular word specified in the beginning of each trial. The task has been used previously with children (e.g., Tyler, Marslen-Wilson, 1981), and was chosen for this experiment because of the relatively low demands on responding participants.

Participants

The experiment was presented to 17 children (8 boys) between ranging in age from 53 to 72 months, with mean age 64 months. All children participated in a longitudinal study of language and language impairment, but none of them was affected by language impairment (SLI). Sixteen children were taken from the control group of unaffected children, and one was an unaffected sibling of a SLI child. Children were recruited from preschools in the areas of Lawrence, Topeka and west Kansas City metro area in Kansas, USA. Children were referred by their teachers or school speech-language professionals, and participated with their parents' explicit consent.

The children were administered several language and nonverbal cognitive tests: the verbal and nonverbal cognitive performance of all children was in the typical range for their ages (within 1.25 standard deviation from the mean).

Stimuli

The structural priming experiment consisted of 24 trials. The trials were presented within an experimental session with a total of 48 trials, 24 from this experiment and 24 from the experiment on error detection. The trials from the two experiments were randomly interspersed. The auditory stimuli were presented from digital recordings. Each trial consisted of a pair of sentences and was preceded by a verbal instruction specifying the target word for word-monitoring in the corresponding trial.

The experiment involved two conditions, matching structure and non-matching structure, with 12 trials in each condition. The conditions differed in the grammatical structure of the first sentence within each trial. In the matching structure condition, the first sentence had the structure Subject-Modal-Verb-Object, while in the non-matching structure condition the first sentence was without a modal (Subject-Verb-Object). The second sentence always had the Subject-Modal-Verb-Object structure. The second sentence in each trial contained the target word for word-monitoring as its object. Table 1 shows a sample trial in each condition.

Table 1. Example of two trials in the structural component, created from one base trial. The target word is italicized.

Condition	First sentence (prime)	Second sentence (target)
Matching	The woman can fix her car.	The cup would hold the *salt.*
Non-matching	The woman fixes her car.	The cup would hold the *salt.*

The 24 trials in the experiments were based on 12 pairs of sentences. The list of sentence pairs is reprinted in the Appendix. Each of the 12 sentence pairs was used once in each condition. The complete set of sentence pairs was presented twice, once in the first half of the session and once in the second half. Six sentence pairs were first presented in the matching structure condition, and the other 6 in the non-matching structure condition. Each sentence pair was thus presented in the first half of the session in one condition, and in the second half of the session in the opposing condition. The ordering in each half was pseudo-randomized so that trials from the two conditions were interspersed with approximately equal frequency during the whole session. The appendix shows trial numbers for each sentence pair in each condition. The trial number is the sequence number that corresponds to the placement of the trial within the experimental session. The same pseudo-random ordering was used for all children, so that the examiners could easily follow the children's responses during the experimental session.

Procedure

The experiment was presented using a laptop computer. The DMDX software (Forster, Forster, 2003) was used for stimulus presentation and response timing and collection. The experimental session had the form of a computer game, where the goal was to press a button when the required target word was heard. The response button was one of the touchpad buttons below the keyboard of the laptop computer; the button was marked with a red cross. The children were seated in front of the laptop computer, and headphones were placed on their heads. The examiner listened on a second pair of headphones. During the training session, the examiners pointed to the response button when the recorded instructions explained its function, and when needed, they provided additional feedback.

The reaction time was measured from the onset of the target word. After the offset of the target word there was a period of 2500 ms during which the responses were collected. If children did not respond within this time, the trial was marked as no response and the next sentence or trial was presented. The whole experimental session, including training, lasted between 16 and 20 minutes, depending on how long a child spent meeting the criteria of the training session.

Analysis

The dependent variable in the analyses was the reaction time. There were two independent variables and two grouping factors that defined the random effects. One independent variable was the experimental condition, a two-level categorical variable. The other was the trial number, i.e. the sequential position of the trial within the session; this was a continuous covariate in the analysis. Two additional variables were used as grouping factors for the definition of random effects. These effects account for the relationships between a group of related observations, e.g. observations from the same person. One grouping factor was the participant: all models fitted in the subsequent analyses contain a random effect for participants. In some analyses, a random effect for sentence pair was estimated as well.

The analyses were performed using the SAS system and its package PROC MIXED. This software is capable of performing all the analyses reported below. The analyses were also replicated using the software system R (R development core team, 2003), package nlme (Pinheiro et al., 2008). This led to numerically identical estimates of random and fixed effects as well as standard errors. However, the hypothesis testing procedure in nlme is slightly different because it does not use the Satterthwaite approximation. Therefore, the p-values produced by nlme were somewhat different.

RESULTS

The mean reaction time in the matching structure condition was 1017 ms (SD = 415), in the non-matching structure condition 998 ms (SD = 437). The difference was not significant when compared using paired sample t-test (t(16) = 0.50, n.s.). There were no significant differences in the number of valid responses (t(16) = 0.42, n.s.), the mean number of valid responses per condition was 9.76.

However, a scatterplot of the data showed that the reaction times in each condition changed differently during the session. To examine the relationships between experimental conditions and the placement of experimental trials, multilevel analysis was performed with condition and trial number as the predictors of interest. The scatterplot also showed that the data were positively skewed, which is usual with reaction time data. The following section first presents a detailed analysis of raw data, including model diagnostics, and then describes the methods that were used to check whether the analysis results were not affected by the skewness or outliers.

Analysis of Raw Data

The first step in the multilevel modeling analysis was fitting a null model with fixed intercept only, and with a random intercept effect for participants. This model has the following structure:

$$RT_{ij} = \beta_0 + \varepsilon_{ij}$$
$$\beta_0 = \gamma_{00} + \zeta_{0i}$$
$$\zeta_0 \sim N(0,\sigma_0),\ \varepsilon \sim N(0,\sigma_\varepsilon)$$

where i stands for participants and j for trials. The level 1 intercept β_0, participant-specific, is decomposed on level 2 into the population intercept γ_{00} and a participant's random effect ζ_{0i}. Level 1 error ε_{ij} is a random error on each trial in every person.

After fitting the null model, more complex models were fit and compared using the likelihood ratio test. The comparisons for models with different fixed effects used the maximum-likelihood (ML) estimates of log-likelihood, comparisons of models that did not differ in their fixed effect structures were based on restricted maximum likelihood (REML) estimates.

The ML fit of the null model had log-likelihood value -2448.8. The first model with fixed predictors added the experimental condition as a fixed effect. The null model was not significantly worse than the model with condition included (χ^2 = 0.2, df = 1, p = 0.65). The next model added trial number as another predictor. This did not lead to a significantly improved fit, either (χ^2 = 0.2, df = 2, p = 0.90) In the next step, an interaction term was added for the interaction between condition and trial number. This model was significantly better than the model with condition only (χ^2 = 9, df = 3, p = 0.029). Since there were no other independent variables of interest, this was accepted as the final structure for the fixed effects.

Subsequent models elaborated the structure of random effects and residuals. The models were compared with the REML log-likelihood estimate for the final model with subject intercept only, which was -2431.5. A model with a random slope of reaction times on trial number was not significantly better than a model with random intercept only (χ^2 = 4.39, df = 2, p = 0.11). This means that there is no evidence that the relationships between reaction time and trial number vary between participants.

The above models assumed that the residuals ε have normal distribution with standard deviation σε. This assumption means that the residual variance is equal for all participants. An alternative is a model assuming different residual variance for each individual, i.e. with subject specific residual variance. Such a model reflects the possibility that different individuals show different amounts of error variance. To explore this possibility, the model estimating residual variance separately for each participant was fit. The model fit significantly better than the model assuming a common residual distribution (χ^2 = 65.9, df = 16, p < 0.001), indicating that individuals differ significantly in the amount of variability in their responses. The formal structure of the model was:

$$RT_{ijk} = \beta_0 + \beta_1 C_j + \beta_2 T_k + \beta_3 (C_j \times T_k) + \varepsilon_{ijk}$$
$$\beta_0 = \gamma_{00} + \zeta_{0i}$$
$$\beta_1 = \gamma_{10}$$
$$\beta_2 = \gamma_{20}$$
$$\beta_3 = \gamma_{30}$$
$$\zeta 0 \sim N(0,\sigma_0),\ \varepsilon \sim N(0,\sigma_{\varepsilon i})$$

where C is the experimental condition and T the trial number. The index in σεi reflects the fact that residuals are different for each participant.

The above model reflects the fact that the design of the experiment includes repeated measures from each participant. However, the design contains another aspect of repeated measurement. The same set of stimuli is presented to every participant, and it is possible that some stimuli elicit systematically faster reactions than others, e.g. because they contain high-frequency words. If this is the case, then the assumption of normal distribution for the error term ε is violated. The mixed model analysis provides an opportunity to address this issue directly by including a crossed random effect for trials or items (Hoffman, Rovine, 2007). The experiment reported here presented a set of 24 trials, so one possibility would be to include a random effect for individual trials. However, the 24 trials were constructed using 12 sentence pairs, and it can be expected that two stimuli from opposing experimental conditions will be more similar if they are based on the same sentence pair and thus share most of their lexical material, rhythmic and prosodic properties etc. The following analysis thus added a random effect for sentence pairs τ, with an assumed normal distribution. The only change against the previous model is in the level 1 structure:

$$RT_{ijkl} = \beta_0 + \beta_1 C_j + \beta_2 T_k + \beta_3 (C_j \times T_k) + \tau_l + \varepsilon_{ijk}$$

where l is the index for a particular sentence pair.

The model was fit but the inclusion of a random item effect did not lead to an improvement, compared to the model with subject intercept random effect only (χ^2 = 1.9, df = 1, p = 0.17). There is thus no evidence that different sentence pairs in this experiment elicit systematically different reaction times. To interpret the findings, we can use the simpler model without the random effect for sentence pairs as final.

The fixed effect estimates from the final model showed a significant interaction between condition and trial number. However, the model's normalized residuals ranged from -2.71 to 3.80. This suggests the presence of outliers in the data. In order to check whether the fixed effect estimates were affected by the extreme observations, the model was refit without the observations with standardized residual values over 3. This corresponds to the recursive method of outlier detection (cf. Van Selst, Jolicoeur, 1994). This criterion resulted in removal of four observations, all with extremely high values.

The estimates of fixed-effect coefficients from the final model with outliers removed appear in Table 2. The estimates suggest a significant interaction between condition and trial number (γ_{30}). The main effects of condition (γ_{10}) and trial number (γ_{20}) are significant as well. Since the condition coded as 1 was the non-matching structure condition, the negative estimate of the main effect for condition means that the reaction times are faster in the non-matching structure condition in the beginning of the experimental session. With increasing trial number, reaction times decrease in the matching structure condition, as suggested by the significant negative main effect for trial number. However, the interaction suggests that in the non-matching structure condition, the reaction times increase: the difference between the interaction estimate and trial number estimate is positive.

Just as in general linear models, such as ANOVA or regression, mixed model analysis should involve the model diagnostics: one should examine whether the distribution of residuals in the model is normal, and whether the results are not influenced by a small number of extreme observations, outliers. One tool for the diagnostics is quantile-quantile plots of residuals from a model against the theoretical normal distribution. If the residual distribution is normal, the points in such a plot will be distributed linearly and sym-

Table 2. Fixed effect and variance component estimates. There are 17 residuals $\sigma_{\varepsilon i}$, one for each participant. The table shows their range. Satterthwaite DF approximation.

Effect	Estimate	DF	t	p
γ_{00}	1095.58	36.2	15.72	<.001
γ_{10}	-172.15	200	-2.86	0.005
γ_{20}	-4.11	202	-2.78	0.006
γ_{30}	6.44	197	3.05	0.003
σ_0	218.98			
Range $\sigma_{\varepsilon i}$	156.46 to 537.59			

metrically along the diagonal. Figure 1 shows the plots for three of the models analyzed above, all with the final structure of fixed effects. Panel a) shows residuals from the model with common residual distribution, panel b) is for the model with subject-specific residuals. It is apparent that the residual distribution of the model with subject-specific residuals is closer to normal: the points in panel b) are aligned closer to the diagonal. Panel c) captures residual distribution in the final model with extreme observations excluded. It indicates that the exclusion of extreme residuals brought the residual distribution closer to normality. On the other hand, the plot is still not completely linear: the histogram of residuals in panel d) confirms that their distribution is skewed. Because of this violation of assumptions, the inferences made from the model should be further scrutinized.

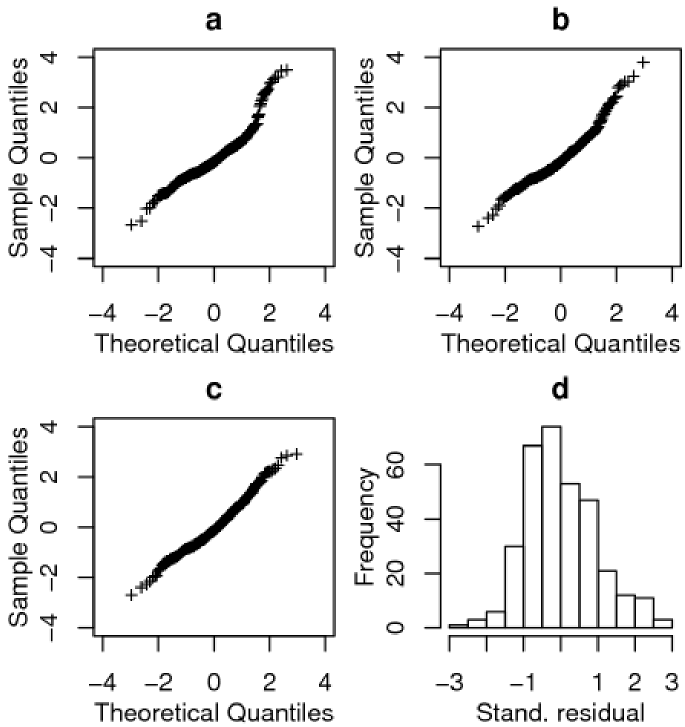

Figure 1. Quantile-quantile plots of residuals from the raw data models with a common residual distribution (a), and subject-specific residual distributions (b). Quantile-quantile plot for the final model with subject-specific residual distributions and removed extreme observations (c), and the histogram of standardized residuals from this model (d).

Analysis of Transformed and Trimmed Data

The skewed distribution of residuals reflects the skewness that typically occurs in reaction time data. There is a large literature about the distributions of reaction time and their statistical models, but there is no generally accepted way to treat or analyze reaction time data. The analyses in this study relied on different transformations and outlier detection methods, and tested whether these data manipulations will change the pattern of results. This is in line with recommendations by Ratcliff (1993), who suggests validating analyses by using multiple cut-off points or transformations. If the analysis of transformed data replicates the pattern of findings from the raw data, it supports the validity of the inferences.

The first re-analysis used logarithmic transformation with natural logarithm because the transformation produced histograms that were visually closer to normality than the original data, and because the logarithmic transformation reflects the multiplicative nature of time measures (Clarke, 1969).

The modeling was driven by the same reasoning as in the analysis of the raw data, and it led to similar decisions. The final model with a random intercept for subjects and subject-specific residuals contained some potentially influential outliers, the normalized residuals ranged from -5.01 to 2.81. The final model was refit without observations with an absolute value of normalized residuals greater than 3, which excluded 5 observations. The estimates from the final model are summarized in Table 3. The pattern of results is similar to that in the analysis of the raw data, even though the parameter estimates converted to the original millisecond metrics are somewhat different. The difference, however, does not change the basic shape of the interaction.

Besides the re-analysis with log-transformed data, further re-analyses were performed with other transformations, as well as with trimmed data. Two transformations were tested, square root and inverse. After both transformations, the analyses arrived at the same final model, as in the above analyses, and the results were similar. The interaction was significant in the resulting models, with reaction times decreasing in the matching structure condition and in-

Table 3. Fixed effect and variance component estimates in the model with log-transformed. Satterthwaite DF approximation.

Effect	Estimate	Converted	DF	t	p
γ_{00}	6.95128	1044.49	38.4	99.08	<.001
γ_{10}	-0.20418	192.91	224	-3.24	0.001
γ_{20}	-0.00373	3.89	227	-2.46	0.015
γ_{30}	0.00784	8.22	225	3.56	<.001
σ_0	0.21914				
Range $\sigma_{\varepsilon i}$	0.167 to 0.474				

creasing in the non-matching structure condition. The estimates suggested that the initial difference between reaction times diminished and changed its direction over the course of the experiment. The same confirmation of the results came from the analyses of trimmed data: one analysis excluded all observations above 2000 ms, the other excluded all observations below and above 2 standard deviations from the overall mean. The findings are thus unlikely to be affected by the skewness or outliers in the data.

DISCUSSION

The analyses found evidence for an interaction between the experimental condition and trial placement. Such an interaction would not be discovered without the inclusion of trial number as a continuous predictor, because the overall means in each condition are similar. Multilevel analysis revealed that although the mean reaction time was similar across conditions, the reaction times changed differently in each condition over the course of the experimental session. The ability of multilevel analysis to incorporate a continuous independent variable thus contributed to an important finding that would be hard to detect using other analytic approaches.

The interaction between trial number and experimental condition may have different explanations. One possibility is that it reflects fatigue and a resulting loss of attention during the experimental session. The participants may become less sensitive to the differences between the experimental conditions as the experiment proceeds, and the initial difference in response latencies may diminish, leading to the opposite slope of latency changes in opposite conditions. In such a situation, the reaction times at the end of the session should be approximately equal at the beginning of the session. This was not the case in the current experiment; the latency differences between the conditions are about as large in the beginning of the session and at the end, but in the opposite directions.

In interpreting the results, two aspects of the findings are of interest. One is the performance of participants in the initial stages of the experiment, the other is the interaction between condition and trial placement. The responses in the early trials of the experiment suggested that the matching structure condition elicited slower responses than the condition with non-matching structures within a sentence pair. This is an unexpected result from the viewpoint of research on syntactic priming: if syntactic priming occurs, exactly the opposite pattern would be expected. A failure to replicate priming effects could lead to a null finding, but there is no reason why it should lead to facilitation in the non-matching structure condition. The most likely explanation is that the initial facilitation of non-matching structure trials is not due to priming but to dishabituation or novelty preference. Novelty preference is often encountered in experiments testing young children's discrimination between two types of stimuli (Jusczyk, 1985; Spelke, 1985). The trials in the current experiment consisted of sentence pairs. The first sentence in a pair appears to work as a habituation stimulus, and the structural difference between the first and second sentence then leads to dishabituation resulting in increased attention and faster response on the final word of the second sentence.

The responses in the initial part of the experiment thus appear to be influenced by a mechanism other than syntactic priming. Syntactic priming, however, is likely to be responsible for the interaction between the experimental conditions. Apart from syntactic priming effects, there is no other reason to expect that the increase or decrease of reaction time should be contingent upon the experimental condition in this study. The results suggest that the effects of syntactic priming appear after repeated exposure to the target structure, and that the effects on sentence processing are cumulative. In the matching structure condition, the increasing strength of priming of the structure with a modal verb leads to a decrease in reaction times, while in the non-matching structure condition, the modal-less prime structures have an increasingly inhibitory effect. Such an effect has not yet been demonstrated in reaction-time studies but is consistent with other findings and theoretical accounts. It has been described previously how repeated presentation of priming stimuli increases the strength of structural priming (Kaschak et al., 2006). This is also consistent with the implicit learning theory of syntactic priming (Chang et al., 2006), according to which the phenomenon is a manifestation of learning mechanisms responsible for the acquisition of structural relationships in language.

To summarize, the results indicate that 5-year-old children discriminate between sentences with different syntactic structures, even in the absence of lexical or semantic overlap between the contrasting sentences. The discrimination appears to be manifested by two separate mechanisms, habituation and syntactic priming.

The study presented here benefited from the use of multilevel modeling analysis in several respects. One was the ability to include a continuous independent variable and evaluate its interactions with the experimental condition. This made it possible to detect an effect of syntactic priming that is not apparent when comparing overall means in each condition. Estimation of subject random effects and subject-specific error variance contributed to an efficient detection of potential outliers. The study demonstrated that multilevel modeling is a flexible tool for the analysis of experimental data.

Received March 19, 2008

REFERENCES

BOCK, K., 1986, Syntactic persistence in language production. *Cognitive Psychology*, 18, 3, 355-387.

BOCK, K., 1989, Closed-class immanence in sentence production. *Cognition*, 31, 2, 163-186.

BOCK, K., LOEBELL, H., 1990, Framing sentences. *Cognition*, 35, 1, 1-39.

BOCK, K., LOEBELL, H., MOREY, R., 1992, From conceptual roles to structural relations: Bridging the syntactic cleft. *Psychological Review*, 99, 1, 150-171.

BRANIGAN, H.P., PICKERING, M.J., CLELAND, A.A., 1999, Syntactic priming in written production: Evidence for rapid decay. *Psychonomic Bulletin & Review*, 6, 4, 635-640.

BRANIGAN, H.P., PICKERING, M.J., STEWART, A.J., McLEAN, J.F., 2000, Syntactic priming in spoken production: Linguistic and temporal interference. *Memory & Cognition*, 28, 8, 1297-1302.

BRYK, A.S., RAUDENBUSH, S.W., 1992, *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage Publications.

CHANG, F., DELL, G.S., BOCK, K., 2006, Becoming syntactic. *Psychological Review*, 113, 2, 234-272.

CHANG, F., DELL, G.S., BOCK, K., GRIFFIN, Z.M., 2000, Structural priming as implicit learning: A comparison of models of sentence production. *Journal of Psycholinguistic Research*, 29, 2, 217-229.

CLARKE, G.M., 1969, *Statistics and experimental design*. London: Edward Arnold Publishers.

COHEN, J., 1983, The cost of dichotomization. *Applied Psychological Measurement*, 7, 3, 249-253.

CORLEY, M., SCHEEPERS, C., 2002, Syntactic priming in English sentence production: Categorical and latency evidence from an Internet-based study. *Psychonomic Bulletin & Review*, 9, 1, 126-131.

ESTIVAL, D., 1985, Syntactic priming of the passive in English. *Text*, 5, 7-22.

FORSTER, K.L., FORSTER, J.C., 2003, DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments and Computers*, 35, 1, 116-124.

FOX TREE, J.E., MEIJER, P.J.A., 1999, Building syntactic structure in speaking. *Journal of Psycholinguistic Research*, 28, 1, 71-92.

HAARMANN, H.J., KOLK, H.H.J., 1994, On-line sensitivity to subject-verb agreement violations in Broca's aphasics: The role of syntactic complexity and time. *Brain and Language*, 46, 4, 493-516.

HARTSUIKER, R.J., KOLK, H.H.J., 1998, Syntactic persistence in Dutch. *Language and Speech*, 41, 2, 143-184.

HOFFMAN, L., ROVINE, M.J., 2007, Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods*, 39, 1, 101-117.

JUSCZYK, P.W., 1985, The high-amplitude sucking technique as a methodological tool in speech perception research. In: N.A. Krasnegor, G. Gottlieb (Eds.), *Measurement of audition and vision in the first year of postnatal life: A methodological overview* (pp. 195-222). Westport, CT: Ablex Publishing.

KASCHAK, M.P., LONEY, R.A., BORREGGINE, K.L., 2006, Recent experience affects the strength of structural priming. *Cognition*, 99, 3, B73-B82.

LEVELT, W.J.M., KELTER, S., 1982, Surface form and memory in question answering. *Cognitive Psychology*, 14, 1, 78-106.

MacCALLUM, R.C., ZHANG, S., PREACHER, K.J., RUCKER, D.D., 2002, On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7, 1, 19-40.

NOPPENEY, U., PRICE, C.J., 2004, An fMRI study of syntactic adaptation. *Journal of Cognitive Neuroscience*, 16, 4, 702-713.

PICKERING, M.J., BRANIGAN, H.P., 1998, The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, 39, 4, 633-651.

PICKERING, M.J., BRANIGAN, H.P., CLELAND, A.A., STEWART, A.J., 2000, Activation of syntactic information during language production. *Journal of Psycholinguistic Research*, 29, 2, 205-216.

PINHEIRO, J.C., BATES, D.M., 2000, *Mixed-effects models in S and S-PLUS.* New York: Springer.

PINHEIRO, J., BATES, D., DebROY, S., SARKAR, D., 2008, R Core team. nlme: Linear and Nonlinear Mixed Effects Models.

POTTER, M.C., LOMBARDI, L., 1998, Syntactic priming in immediate recall of sentences. *Journal of Memory and Language*, 38, 3, 265-282.

R development core team, 2003, R (programmable environment for statistical computing). Retrieved from http://cran.r-project.org.

RATCLIFF, R., 1993, Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114, 3, 510-532.

SCHEEPERS, C., 2003, Syntactic priming of relative clause attachments: Persistence of structural configuration in sentence production. *Cognition*, 89, 3, 179-205.

SMITH, M., WHEELDON, L., 2001, Syntactic priming in spoken sentence production - an online study. *Cognition*, 78, 2, 123-164.

SPELKE, E.S., 1985, Preferential-looking methods as tools for the study of cognition in infancy. In: N.A. Krasnegor, G. Gottlieb (Eds.), *Measurement of audition and vision in the first year of postnatal life: A methodological overview* (pp. 323-363). Westport, CT: Ablex Publishing.

TYLER, L.K., MARSLEN-WILSON, W.D., 1981, Children's processing of spoken language. *Journal of Verbal Learning and Verbal Behavior*, 20, 4, 400-416.

VAN SELST, M., JOLICOEUR, P., 1994, A solution to the effect of sample size on outlier elimination. *Quarterly Journal of Experimental Psychology*, A 47, 3, 631-650.

APPENDIX

Sentence pairs used as stimuli, along with the numbers of trials that used the sentence pair in the matching (first number) and non-matching (second number) structure condition.

1. The woman can fix her car. The cup would hold the salt. 9 29
2. The teacher could close the book. The grandma can write her name. 27 10
3. Superman would watch the train. The friend will call her home. 7 32

Appendix continues

Appendix (continued)

4. The cat will use her nose. The dad could wear the pants. 30 11
5. The finger can touch the wire. The cow will jump the hole. 6 36
6. The pig will push the door. The baby would see her shoe. 31 3
7. The boat would catch the fish. The bear could like his den. 24 38
8. The machine could make the food. The bird can pick the rice. 48 22
9. The fire can burn the shirt. The guy will paint the wall. 18 43
10. The soap could clean the toy. The bus would carry the bag. 41 13
11. The hand will drop the fork. The horse could eat the corn. 19 45
12. The mouse would taste the cheese. The girl can swing the stick. 44 17

POUŽITÍ SMÍŠENÝCH REGRESNÍCH MODELŮ PRO ANALÝZU EXPERIMENTÁLNÍCH DAT: KUMULATIVNÍ EFEKTY STRUKTURNÍHO PRIMINGU U DĚTÍ

F. S m o l í k

Souhrn: Smíšené regresní modely (známé též jako víceúrovňové modely) představují flexibilní alternativu ke tradiční analýze opírající se o faktoriální analýzu rozptylu pro opakovaná měření. Tento článek popisuje psycholingvistický experiment a detailně vysvětluje postup analýzy dat získaných v tomto experimentu, přičemž demonstruje využití smíšených modelů při analýze experimentu s kategoriálními a kontinuálními prediktory. Demonstruje rovněž zahrnutí komplexních předpokladů o vlastnostech chybové variance v experimentu do analýzy. Samotný experiment zkoumal efekty strukturního primingu na reakční čas v úloze monitorování slov. Čtyř- a pětileté děti slyšely dvojice vět, jejichž gramatické struktury se shodovaly nebo byly rozdílné. Děti měly za úkol stisknout tlačítko po identifikaci předem zadaného slova, které se objevovalo na konci druhé věty. Analýza pomocí smíšených modelů ukázala statisticky významnou interakci mezi experimentální podmínkou a pozicí položky v experimentu. V položkách se shodnými strukturami docházelo k postupnému poklesu reakčních časů, zatímco v položkách s rozdílnými strukturami vět se reakční časy prodlužovaly. Analýza demonstruje, jak lze smíšených modelů použít k identifikaci změny v odpovědích pokusných osob v průběhu experimentu.