A Markov Mixed-Effect Multinomial Logistic Regression Model for Nominal Repeated Measures with an Application to Syntactic Self-Priming Effects

Sun-Joo Cho, Vanderbilt University
Duane Watson, Vanderbilt University
Cassandra Jacobs, University of Wisconsin, Madison
Matthew Naveiras, Vanderbilt University

February 17, 2020

Accepted for publication in *Multivariate Behavioral Research*

A Markov Mixed-Effect Multinomial Logistic Regression Model for Nominal Repeated Measures with an Application to Syntactic Self-Priming Effects

Abstract

Syntactic priming effects have been investigated for several decades in psycholinguistics and the cognitive sciences to understand the cognitive mechanisms that support language production and comprehension. The question of whether speakers prime themselves is central to adjudicating between two theories of syntactic priming, activation-based theories and expectation-based theories. However, there is a lack of a statistical model to investigate the two different theories when nominal repeated measures are obtained from multiple participants and items. This paper presents a Markov mixed-effect multinomial logistic regression model in which there are fixed and random effects for own-category lags and cross-category lags in a multivariate structure and there are category-specific crossed random effects (random person and item effects). The model is illustrated with experimental data that investigates the average and participant-specific deviations in syntactic self-priming effects. Results of the model suggest that evidence of self-priming is consistent with the predictions of activation-based theories. Accuracy of parameter estimates and precision is evaluated via a simulation study using Bayesian analysis.

Index Terms: Bayesian analysis, crossed random effects, generalized linear mixed effect model, lag effects, Markov model, multinomial logistic regression model, psycholinguistics

Introduction

Study Motivation: Syntactic Priming Effects

This study is inspired by investigations of syntactic priming effects in psycholinguistics. Syntactic priming describes a phenomenon in which speakers are more likely to use a grammatical structure in the future if they have been exposed to that grammatical structure in the past (Bock, 1986; Mahowald, James, Futrell, & Gibson, 2016). For example, in describing a gift-giving event at a party, an English speaker might say either, "Eleanor gave Penelope a book" (double object [DO] structure) or "Penelope gave a book to Eleanor" (prepositional object [PO] structure). Although the decision to choose one of these sentences is influenced by a host of different factors, one factor is whether or not the speaker has heard a DO structure or a PO structure earlier in the conversation (e.g., "Eleanor handed Penelope a drink" vs. "Eleanor handed a drink to Penelope"). The goal of this paper is to propose a statistical model to provide substantive researchers in psycholinguistics and the cognitive sciences with a new model-based perspective of the syntactic priming effect.

For several decades, a core research question in psycholinguistics and the cognitive sciences has been understanding whether syntactic priming occurs. The question of the extent to which previous exposure to a linguistic phenomenon influences speakers' linguistic choices is widespread in the psycholinguistics literature. Studies have investigated whether the ways in which speech sounds (Giles, Coupland, & Coupland, 1991), word choice (Brennan & Clark, 1996), and conceptual representations (Gruberg, Ostrand, Momma, & Ferreira, 2019) are influenced by previous encounters with similar linguistic information. Research on syntactic priming has helped to shed theoretical light on the cognitive mechanisms that support language production and comprehension (Bock, 1986; Chang, Dell, & Bock, 2006; Pickering & Garrod, 2004; Pickering & Garrod, 2013; Reitter, Keller, & Moore, 2011; and many others). While it is clear that exposure to a syntactic structure can increase a speakers' tendency towards using that structure in the future, it is likely not the only influence. In particular, whether speakers can prime themselves or not (i.e., *syntactic self-priming*), can reveal the nature of the underlying representations of the language system, outlined below.

There are two broad classes of theories of syntactic priming: activation-based theories and expectation-based theories. Expectation-based theories proposes that through a lifetime of experience with language, speakers build a model of grammatical frequency, which is used in both making predictions in language comprehension, and choosing structures in language production (e.g., Chang et al., 2006; MacDonald, 2013). In activated-based theories, grammatical structures vary in baseline activation, and speakers prefer structures that are more activated over structures that are less activated (e.g., Reitter et al., 2011). Critically, activation based theories predict that speakers' own grammatical choices should influence speakers' choices in the future. In contrast, expectation-based theories predict that a speaker's grammatical choices should not influence future productions because a speaker's own productions cannot be unexpected. Thus, the question of whether speakers prime themselves is central to adjudicating between these two fundamentally different classes of theories of how the language production system works.

Despite being of great theoretical interest to linguistics and cognitive psychology, the analytical tools that have been used to investigate syntactic self-priming do not directly address whether there are dependencies between productions. In this domain of syntactic priming, only a handful of studies have investigated self-priming (Gries, 2005; Reitter & Moore, 2014). These have looked at factors that predict repetition of a grammatical structure (Reitter & Moore, 2014) or used a Chi-square test to determine whether syntactic repetition occurs more often than one would expect by chance (Gries, 2005). Gries (2005) further used a generalized linear model to investigate whether the production of a PO or DO in a corpus of text predicted which sentence structures would be produced in subsequent sentences. However, this approach did not allow for multinomial responses, did not control for spurious correlations between pairs of productions due to overall speaker or discourse biases for one construction over the other, and did not model individual differences in self-priming.

The usefulness of an appropriate method of assessing self-priming in syntactic production has far-reaching advantages for other domains, including studying the perpetuation of response bias in motor tasks in humans and other animals, in which hysteresis, or the repetition of previous states, is often observed (Cohen & Rosenbaum, 2004; Weiss & Wark, 2009; MacDonald & Weiss, 2017). Other areas of language production as well, especially foreign language learning and behavior in artificial grammar tasks, have demonstrated a tendency among speakers to perpetuate their own productions (e.g., Baese-Berk, 2019; Stave, Smolek, & Kapatsinski, 2013). Properly characterizing dependencies between adjacent outcomes is therefore an important tool for broad audiences in cognitive science.

Markov Regression Models, Mixed-Effects Multinomial Logistic Regression Models, and Their Limitations

The dependencies between discrete response categories (e.g., PO and DO) over equally-spaced discrete periods of time or trials can be investigated using an observed Markov regression model[1] in which the conditional means and variances given the past are explicit functions of past outcomes (Cox, 1981). The lag effects to model the dependencies in the observed Markov regression model can be used to assess a transition from one response category to another (e.g., PO $\rightarrow$ DO; DO $\rightarrow$ PO) which could serve as evidence for syntactic self-priming. Although the observed Markov regression model is particularly suited to answer questions regarding syntactic self-priming, the prior literature in this field has not applied this model.

There is a long history of interest in Markov regression modeling in repeated measures (e.g., Azzalini, 1994), panel data (e.g., Korn & Whittemore, 1979) or time-series data (e.g., Cox, 1970). Whereas much of this interest has focused on the Markov models for ordinal data, little attention has been given to nominal responses. One exception is work by Fokianos and Kedem (2003). They presented a Markov multinomial logistic regression model for nominal time series

[1] We use the term 'observed' Markov regression model instead of the latent or hidden Markov regression model because dependence models are obtained using autoregression for observations (i.e., observed nominal responses in our application).

and illustrated the model using DNA sequence data. However, the authors did not consider incorporating dependencies in the data (e.g., person and item clusters), nor variability in lag parameters.

Hedeker (2003, 2008) and Skrondal and Rabe-Hesketh (2003) presented a mixed-effects multinomial logistic regression model to incorporate dependency due to clusters. Although this model is a general model and has been widely applied in many disciplines, Markov modeling was not considered in the applications in Hedeker (2003, 2008) and Skrondal and Rabe-Hesketh (2003). In nominal repeated measures, specification of lag effects can be complicated because there are own-category lags (e.g., DO–DO, PO–PO) and cross-category lags (e.g., DO–PO, PO–DO) in a multivariate structure. The cross-category lag effects between variables have been considered mainly for multivariate linear models (e.g., Bringmann et al., 2017; Lodewyckx, Tuerlinckx, Kuppens, Allen, & Sheeber, 2011). Pettitt, Tran, Haynes, and Hay (2006) presented a *Markov mixed-effect multinomial logistic regression model* for nominal responses (employed, unemployed, or non-participant) to assess the transition from one employment state to another using the Longitudinal Survey of Immigrants to Australia (LSIA) data. In psycholinguistics, simultaneously modeling person and item dependencies with random person effect and random item effect is widely advocated in generalized linear mixed effect models (GLMM) (Baayen, Davidson, & Bates, 2008; Jaeger, 2008). The random item effect is considered in the GLMM to model the fact that items are sampled from an item population (Clark, 1973). However, the model specification in Pettitt et al. (2006) did not consider all possible dependencies in the person-by-item data (e.g., item effects) to test whether individuals remained in the same state (employed, unemployed, or non-participant).

Study Purpose

In this paper, we extend the approach of Pettitt et al. (2006) by including category-specific random person and random item effects in the person-by-item data, and random slopes for lag effects in a Markov mixed-effect multinomial logistic regression model. The aim of this paper is to specify the Markov mixed-effect multinomial logistic regression model accounting for all possible dependencies and variability in the data. Ignoring the dependencies and variability in the data affects the precision of estimates of model parameters and consequently may lead to inaccurate statistical tests for the fixed effects (e.g., the fixed lag effects). The model is illustrated using an experimental data set that tests for a syntactic self-priming effect. The goal of the analysis is to investigate the transition from one category (DO or PO) to another between adjacent trials as evidence of syntactic self-priming, which are modelled with lag covariates in a mixed-effect multinomial logistic regression model. We chose `OpenBUGS`, an open-source program for Bayesian model estimation (Lunn, Spiegelhalter, Thomas & Best, 2009), to fit the model. Software like `OpenBUGS` provides applied researchers with generic and flexible tools for Bayesian analysis. Accuracy of parameter estimates and precision (i.e., posterior standard deviation) was evaluated using `OpenBUGS` for Bayesian analysis with non-informative priors.

The remainder of this paper is organized as follows. In Section 2, we present the Markov mixed-effect multinomial logistic regression model, provide the estimation method using `OpenBUGS`, and describe the model evaluation and testing. In Section 3, the model is illustrated using an empirical data set. In Section 4, parameter recovery of the model is evaluated via a simulation study. In Section 5, we end with a summary and a discussion.

A Markov Mixed-Effect Multinomial Logistic Regression Model for Syntactic Self-Priming

In this section, we present (a) a Markov mixed-effect multinomial logistic regression model to investigate the syntactic self-priming effect, (b) its Bayesian estimation using `OpenBUGS`, and (c) Bayesian model selection, evaluation, and testing.

Model Specification

In investigating the syntactic self-priming effect, the outcome variable had 3 categories, PO, DO and Other. The three outcomes (PO, DO and Other) were considered nominal (or unordered) responses. To avoid the arbitrariness resulting from integer assignment to nominal categories, Other was set as a baseline outcome. Two contrasts (Other vs. DO; Other vs. PO) were modelled. Let y_{tji} be the outcome for trial t ($t = 1, \ldots, T$), person j ($j = 1, \ldots, J$), and item i ($i = 1, \ldots, I$), coded as $y_{tji} = 1$ for Other, $y_{tji} = 2$ for DO, and $y_{tji} = 3$ for PO.

For the first trial ($t = 1$), lag effects do not exist. Thus, separate models for $t = 1$ and $t > 1$ were considered, as used in Heckman (1981). In each model, we consider category-specific random item effects (instead of fixed item effects), because items are assumed to be sampled from an item population, as is often assumed in psycholinguistics (Clark, 1973). In addition, category-specific random person effects were considered in the models, assuming that each response category is related to an underlying latent 'response tendency' for that category (Bock, 1970). Furthermore, random lag slope effects were considered to model individual differences in lag effects, as in multilevel dynamic factor models (e.g., Song & Zhang, 2014) or mixed Markov models (e.g., de Haan-Rietdijk et al., 2017) or multilevel threshold autoregressive models (e.g., de Haan-Rietdijk, Gottman, Bergeman, & Hamaker, 2016).

Because we considered person and item effects in the Markov mixed-effect nominal response model, the model is related to the nominal categories model (Bock, 1972). The Markov mixed-effect nominal response model is different from Bock's (1972) original nominal categories model in the following aspects: (a) the same slope across items is used, (b) the category-specific item parameters are random, as advocated in Baayen et al. (2008), (c) category-specific random person parameters are considered (instead of one random person parameter across categories), and (d) the fixed and random lag effects are modelled.

In this study, the assumptions for dimensionality (i.e., one random person effect across categories vs. category-specific random person effects) and individual differences in random lag slope effects were investigated by comparing models with differing effects. Below, we present the most complex model having category-specific random person and item effects, and random lag slope effects.

A model for $t = 1$. A model for $t = 1$ is written as:

$$logit \begin{pmatrix} P_{1ji2} \\ P_{1ji3} \end{pmatrix} = \begin{pmatrix} log\frac{P(y_{1ji}=2|\gamma_{12},\theta_{1j2},\beta_{1i2})}{P(y_{1ji}=1|\gamma_{12},\theta_{1j2},\beta_{1i2})} \\ log\frac{P(y_{1ji}=3|\gamma_{13},\theta_{1j3},\beta_{1i3})}{P(y_{1ji}=1|\gamma_{13},\theta_{1j3},\beta_{1i3})} \end{pmatrix} = \begin{pmatrix} \eta_{1ji2} \\ \eta_{1ji3} \end{pmatrix}. \quad (1)$$

The linear predictor, η_{1ji2}, is for the logarithm of the probability of DO relative to Other and the linear predictor, η_{1ji3}, is for the logarithm of the probability of PO relative to Other.

The linear predictors are defined as

$$\eta_{1ji2} = \gamma_{12} + \theta_{1j2} + \beta_{1i2}$$

and

$$\eta_{1ji3} = \gamma_{13} + \theta_{1j3} + \beta_{1i3}, \quad (2)$$

where γ_{12} is the grand mean across all observations for Other vs. DO, γ_{13} is the grand mean across all observations for Other vs. PO, θ_{1j2} is a random person intercept for Other vs. DO, θ_{1j3} is a random person intercept for Other vs. PO, β_{1i2} is a random item intercept for Other vs. DO, and β_{1i3} is a random item intercept for Other vs. PO.

A model for $t > 1$. In the current analysis, the goal is to find dependencies between adjacent trials, which were modelled with first-order lag covariates. Because of the multivariate outcomes, two lag covariates (Other vs. DO; Other vs. PO) were created based on two dummy coded variables, C_{tj2} and C_{tj3} (also see Table 1, top):

$$C_{tj2} = \begin{cases} 1 & \text{if } y_{tji} = 2 \\ 0 & \text{otherwise} \end{cases}$$

and

$$C_{tj3} = \begin{cases} 1 & \text{if } y_{tji} = 3 \\ 0 & \text{otherwise.} \end{cases}$$

Using C_{tj2} and C_{tj3}, the first-order lag covariates ($C_{(t-1)j2}$ and $C_{(t-1)j3}$) were created as values at trial $t-1$ (also see Table 1, below for three persons as an example).

=========================

Insert Table 1 about here

==========================

For $t > 1$, the model is written as

$$logit \begin{pmatrix} P_{tji2} \\ P_{tji3} \end{pmatrix} = \begin{pmatrix} log \frac{P(y_{tji}=2|C_{(t-1)j2}, C_{(t-1)j3}, \gamma_2, \lambda_{j22}, \lambda_{j32}, \theta_{j2}, \beta_{i2})}{P(y_{tji}=1|C_{(t-1)j2}, C_{(t-1)j3}, \gamma_2, \lambda_{j22}, \lambda_{j32}, \theta_{j2}, \beta_{i2})} \\ log \frac{P(y_{tji}=3|C_{(t-1)j2}, C_{(t-1)j3}, \gamma_3, \lambda_{j23}, \lambda_{j33}, \theta_{j3}, \beta_{i3})}{P(y_{tji}=1|C_{(t-1)j2}, C_{(t-1)j3}, \gamma_3, \lambda_{j23}, \lambda_{j33}, \theta_{j3}, \beta_{i3})} \end{pmatrix} = \begin{pmatrix} \eta_{tji2} \\ \eta_{tji3} \end{pmatrix}. \tag{3}$$

As in the model $t = 1$, the linear predictor, η_{tji2}, is for the logarithm of the probability of DO relative to Other and the linear predictor, η_{tji3}, is for the logarithm of the probability of PO relative to Other.

The linear predictors are defined as

$$\eta_{tji2} = \gamma_2 + C'_{(t-1)j2}\lambda_{j22} + C'_{(t-1)j3}\lambda_{j32} + \theta_{j2} + \beta_{i2}$$

and

$$\eta_{tji3} = \gamma_3 + C'_{(t-1)j2}\lambda_{j23} + C'_{(t-1)j3}\lambda_{j33} + \theta_{j3} + \beta_{i3},$$

where γ_2 is a mean level across observations of Other for Other vs. DO, γ_3 is a mean level across observations of Other for Other vs. PO, $\boldsymbol{\lambda}_j = [\lambda_{j22}, \lambda_{j32}, \lambda_{j23}, \lambda_{j33}]'$ is the lag effects (which is defined in detail below), θ_{j2} is a random person intercept for Other vs. DO, θ_{j3} is a random person intercept for Other vs. PO, β_{i2} is a random item intercept for Other vs. DO, and β_{i3} is a random item intercept for Other vs. PO.

First-order lag parameter. In this subsection, our primary interest, the lag effects ($\boldsymbol{\lambda}_j = [\lambda_{j22}, \lambda_{j32}, \lambda_{j23}, \lambda_{j33}]'$) are explained:

- λ_{j22} is the lag effect for the current response Other vs. DO due to the previous response changing from Other to DO.
- λ_{j32} is the lag effect for the current response Other vs. DO due to the previous response changing from Other to PO.

- λ_{j23} is the lag effect for the current response Other vs. PO due to the previous response changing from Other to DO.

- λ_{j33} is the lag effect for the current response Other vs. PO due to the previous response changing from Other to PO.

In estimation, $\boldsymbol{\lambda}_j = [\lambda_{j22}, \lambda_{j32}, \lambda_{j23}, \lambda_{j33}]'$ is decomposed as follows:

- $\lambda_{j22} = \lambda_{22} + \lambda_{R.j22}$, where λ_{22} is the fixed lag effect and $\lambda_{R.j22}$ is the random lag slope effect (person-specific deviation) from λ_{22}.

- $\lambda_{j32} = \lambda_{32} + \lambda_{R.j32}$, where λ_{32} is the fixed lag effect and $\lambda_{R.j32}$ is the random lag slope effect (person-specific deviation) from λ_{32}.

- $\lambda_{j23} = \lambda_{23} + \lambda_{R.j23}$, where λ_{23} is the fixed lag effect and $\lambda_{R.j23}$ is the random lag slope effect (person-specific deviation) from λ_{23}.

- $\lambda_{j33} = \lambda_{33} + \lambda_{R.j33}$, where λ_{33} is the fixed lag effect and $\lambda_{R.j33}$ is the random lag slope effect (person-specific deviation) from λ_{33}.

Figure 1 presents a graphical representation for $\boldsymbol{\lambda}_j$. We focus on interpreting effects of own-category lags (λ_{j22} for Other vs. DO [$y_{tji} = 1$ vs. $y_{tji} = 2$] and λ_{j33} for Other vs. PO [$y_{tji} = 1$ vs. $y_{tji} = 3$], as presented with lines in Figure 1). That is, the estimates of λ_{j22} and λ_{j33} will inform us of DO-DO and PO-PO reuse, which will potentially serve as evidence of syntactic self-priming. The effects of cross-category lags (λ_{j32} and λ_{j23}) were added in the model as controlling covariates for Other vs. DO [$y_{tji} = 1$ vs. $y_{tji} = 2$] and for Other vs. PO [$y_{tji} = 1$ vs. $y_{tji} = 3$], respectively (as presented with dotted lines in Figure 1).

==========================

Insert Figure 1 about here

==========================

A joint model for $t = 1, \dots, T$. A joint model for all trials ($t = 1, \dots, T$) can be presented by introducing an indicator function $I[\cdot]$ as follows:

$$\zeta_{tji2} = \eta_{1ji2} I[t = 1] + \eta_{tji2} I[t > 1]$$

$$= (\gamma_{12} + \theta_{1j2} + \beta_{1i2}) I[t = 1] + (\gamma_2 + C'_{(t-1)j2}\lambda_{j22} + C'_{(t-1)j3}\lambda_{j32} + \theta_{j2} + \beta_{i2}) I[t > 1] \tag{4}$$

and

$$\zeta_{tji3} = \eta_{1ji3} I[t = 1] + \eta_{tji3} I[t > 1]$$

$$= (\gamma_{13} + \theta_{1j3} + \beta_{1i3}) I[t = 1] + (\gamma_3 + C'_{(t-1)j2}\lambda_{j23} + C'_{(t-1)j3}\lambda_{j33} + \theta_{j3} + \beta_{i3}) I[t > 1]. \tag{5}$$

To identify the model, the linear predictor for the baseline outcome (Other), $\zeta_{tji1} = 0$, is set, which implies that all parameters in the linear predictor ζ_{tji1} are set to 0. In addition, the means of random effects in the linear predictors ζ_{tji2} and ζ_{tji3} are set to 0. For $t = 1$, two random person effects and two random item effects are assumed to follow a multivariate normal distribution, respectively: $\boldsymbol{\theta}_{1j} = [\theta_{1j2}, \theta_{1j3}]' \sim MN(\mathbf{0}, \Sigma_1)$ and $\boldsymbol{\beta}_{1i} = [\beta_{1i2}, \beta_{1i3}]' \sim MN(\mathbf{0}, \Sigma_2)$. For $t > 1$, six random person effects and two random item effects are assumed to follow a multivariate normal distribution, respectively: $\boldsymbol{\theta}_j = [\theta_{j2}, \theta_{j3}, \lambda_{R.j22}, \lambda_{R.j32}, \lambda_{R.j23}, \lambda_{R.j33}]' \sim MN(\mathbf{0}, \Sigma_3)$ and $\boldsymbol{\beta}_i = [\beta_{i2}, \beta_{i3}]' \sim MN(\mathbf{0}, \Sigma_4)$.

The probability of having each category score k ($k = 1, 2, 3$), $P(y_{tji} = k | C_{(t-1)j2}, C_{(t-1)j3}, \boldsymbol{\gamma}, \boldsymbol{\theta}_{1j}, \boldsymbol{\beta}_{1i}, \boldsymbol{\lambda}_j, \boldsymbol{\theta}_j, \boldsymbol{\beta}_i)$, can be specified as follows:

$$P(y_{tji} = 1 | C_{(t-1)j2}, C_{(t-1)j3}, \boldsymbol{\gamma}, \boldsymbol{\theta}_{1j}, \boldsymbol{\beta}_{1i}, \boldsymbol{\lambda}_j, \boldsymbol{\theta}_j, \boldsymbol{\beta}_i) = \frac{1}{1 + exp(\zeta_{tji2}) + exp(\zeta_{tji3})}, \tag{6}$$

$$P(y_{tji} = 2 | C_{(t-1)j2}, C_{(t-1)j3}, \boldsymbol{\gamma}, \boldsymbol{\theta}_{1j}, \boldsymbol{\beta}_{1i}, \boldsymbol{\lambda}_j, \boldsymbol{\theta}_j, \boldsymbol{\beta}_i) = \frac{exp(\zeta_{tji2})}{1 + exp(\zeta_{tji2}) + exp(\zeta_{tji3})}, \tag{7}$$

and

$$P(y_{tji} = 3 | C_{(t-1)j2}, C_{(t-1)j3}, \boldsymbol{\gamma}, \boldsymbol{\theta}_{1j}, \boldsymbol{\beta}_{1i}, \boldsymbol{\lambda}_j, \boldsymbol{\theta}_j, \boldsymbol{\beta}_i) = \frac{exp(\zeta_{tji3})}{1 + exp(\zeta_{tji2}) + exp(\zeta_{tji3})}, \tag{8}$$

where $\boldsymbol{\gamma} = [\gamma_{12}, \gamma_{13}, \gamma_2, \gamma_3]'$ and $exp(\zeta_{tji1}) = exp(0) = 1$.

Parameter Estimation

In the model, random person parameters ($\boldsymbol{\theta}_1, \boldsymbol{\theta}$) and random item parameters ($\boldsymbol{\beta}_1, \boldsymbol{\beta}$) are crossed random effects. In psycholinguistics, GLMM with crossed random effects (i.e., random person effects and random item effects) have become the dominant analysis method because they can account for dependencies attributable to persons or items (Baayen et al., 2008). However, commonly used R packages for models with the crossed random effects such as `lme4` (Bates et al., 2018) are not available for multinomial models. `MCMCglmm` (Hadfield, 2010) can be used for multinomial models, but it is not feasible for the category-specific effects we consider in the illustration. In this study, we chose to implement Bayesian analysis in `OpenBUGS` to avoid high-dimensional integration in the model with crossed random effects. Below, the likelihood function, prior and hyper-prior distributions, and a joint posterior distribution are specified for the Markov mixed-effect multinomial logistic regression model. The `OpenBUGS` code is presented in Appendix A.

The likelihood function for the model can be defined as

$$f(\mathbf{y}|\Theta) = \prod_{t=1}^{T}\prod_{j=1}^{J}\prod_{i=1}^{I}\prod_{k=1}^{K} P(y_{tji} = k|C_{(t-1)j2}, C_{(t-1)j3}, \boldsymbol{\gamma}, \boldsymbol{\theta}_{1j}, \boldsymbol{\beta}_{1i}, \boldsymbol{\lambda}_j, \boldsymbol{\theta}_j, \boldsymbol{\beta}_i)^{[y_{tji}=k]}, \tag{9}$$

where $\Theta = [\gamma_{1k}, \gamma_k, \lambda_{2k}, \lambda_{3k}, \boldsymbol{\theta}_1, \boldsymbol{\beta}_1, \boldsymbol{\theta}, \boldsymbol{\beta}, \Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4]$ for $k = 2, 3$. The joint posterior distribution for the parameters Θ is written as

$$P(\Theta|\mathbf{y}) \propto f(\mathbf{y}|\Theta) \cdot [P(\gamma_{1k})P(\gamma_k)P(\lambda_{2k})P(\lambda_{3k})P(\boldsymbol{\theta}_{1j}|\mathbf{0}, \Sigma_1)P(\boldsymbol{\beta}_{1i}|\mathbf{0}, \Sigma_2)P(\boldsymbol{\theta}_j|\mathbf{0}, \Sigma_3)P(\boldsymbol{\beta}_i|\mathbf{0}, \Sigma_4)]$$

$$\cdot [P(\Sigma_1)P(\Sigma_2)P(\Sigma_3)P(\Sigma_4)], \tag{10}$$

where probabilities within the first bracket indicate prior distributions and probabilities within the second bracket indicate hyper-prior distributions. All parameters in the joint model for $t = 1, \ldots, T$ were sampled from the joint posterior distribution after setting $C_{0j2} = 0$ and $C_{0j3} = 0$ at $t = 1$. A long form of the data (as shown in Table 1) was used in estimation instead of a wide form because the long form is an elegant way of dealing with missing values from the *random* assignment of items to persons and trials in our illustrative data.

Because we do not have relevant prior information on parameters available, non-informative priors and hyper-priors were considered. Prior and hyper-prior distributions were specified in `OpenBUGS` as follows:

$$\begin{aligned}
\gamma_{1k} &\sim N(0, 0.001), \\
\gamma_{k} &\sim N(0, 0.001), \\
\lambda_{2k} &\sim N(0, 0.001), \\
\lambda_{3k} &\sim N(0, 0.001), \\
\boldsymbol{\theta}_{1j} &\sim MN(\mathbf{0}, \Sigma_{1(2\times 2)}) \\
\boldsymbol{\beta}_{1i} &\sim MN(\mathbf{0}, \Sigma_{2(2\times 2)}) \\
\boldsymbol{\theta}_{j} &\sim MN(\mathbf{0}, \Sigma_{3(6\times 6)}) \\
\boldsymbol{\beta}_{i} &\sim MN(\mathbf{0}, \Sigma_{4(2\times 2)}) \\
\Sigma_{1(2\times 2)} &\sim Wishart(I_2, 3), \\
\Sigma_{2(2\times 2)} &\sim Wishart(I_2, 3), \\
\Sigma_{3(6\times 6)} &\sim Wishart(I_6, 7),
\end{aligned}$$

and

$$\Sigma_{4(2\times 2)} \sim Wishart(I_2, 3),$$

where $k = 2, 3$. Non-informative priors on fixed parameters were imposed (variance=1/0.001=1000 where 1,000 is the variance in `OpenBUGS`). Inverse-Wishart hyper-priors on variance and covariance matrices ($Wishart$ in `OpenBUGS`) were chosen as conjugate priors for multivariate normal distributed random effects. For the non-informative inverse-Wishart prior, commonly used parameters of the distribution is the unit matrix I of size D (I_D, where D is the rank of the random effects) and a small degree of freedom (df) as D or $D+1$. Having $df = D$ indicates that the prior is minimally informative (e.g., Gelman & Hill, 2007). We set $df = D + 1$, which implies that each of the correlations in the variance and covariance matrix has a uniform prior distribution marginally (Gelman, Carlin, Stern, Dunson, & Rubin, 2013).

In this study, the non-informative inverse-Wishart prior distribution on variance and covariance matrix was chosen as a conjugate prior. It has been found that such a non-informative prior specification can be informative when the variance of random effects is small (e.g., < 0.0225 in Schuurman et al. [2016]). Schuurman, Grasman, and Hamaker (2016) compared three different priors (Inverse-Wishart distributions with hyper-priors, maximum likelihood [ML] estimates, and default conjugate priors) on the variance and covariance matrix for the multivariate linear mixed autoregressive model (having three random effects). They found that the data-based ML prior specification outperformed the other two priors in the simulation conditions that have small variances of random lag slopes (0.01, 0.0225, 0.025) and small sample sizes (25, 50, 75). Following findings from Schuurman et al. (2016), the results using the data-based ML priors were compared with those of the inverse-Wishart distribution for the variance and covariance matrix ($Wishart(I_D, D+1)$ in `OpenBUGS`). Because the ML estimates of the Markov mixed-effect multinomial logistic regression model (Equations 4 and 5) are not available in currently available software, the ML estimates from simpler models that allow ML estimation were considered to generate preliminary results for the variance and covariance matrix in the Markov mixed-effect multinomial logistic regression model. The simultaneous estimation of two separate models at trial 1 and trials 2-T, and the crossed random effects (random person and item effects) lead to estimation challenges. Thus, the simpler models we chose were two separate analyses of mixed multinomial logistic regression models at trial 1 and trials 2-T respectively, with the random person and the fixed item effects. The ML estimates of the variance and covariance matrix of random person effects in the simpler models and sample variance and covariance matrix of fixed item parameter estimates were used as the data-based priors, and estimates were obtained in Mplus version 8.1 (Muthén & Muthén, 1998–2017).

For convergence diagnostic, the potential scale reduction factor (PSRF; Gelman & Rubin,

1992) was considered with three chains, and the PSRF value of 1.01 was used as a threshold to indicate model convergence (Gelman et el., 2013). In addition, autocorrelation plots were examined to ensure that Markov chain Monte Carlo (MCMC) samples can be summarized using descriptive statistics (e.g., posterior mean and posterior standard deviation) like conventional independent samples. The lower the autocorrelation, the greater the amount of information contained in a given number of samples from the posterior (called efficiency). In the presence of autocorrelation, the samples from MCMC can be thinned by keeping every hth draw to produce samples that are more nearly independent, and discarding the rest. The *thinned* samples reduce the autocorrelation in the samples. However, this thinning reduces the total number of samples after convergence (denoted by L) with which posterior descriptive statistics are calculated. It has been noted that the average over thinned samples has greater variance than the plain average over all of the samples (Geyer, 1992). That is, thinning can never be as efficient as using all the iterations. Thus, when we observe the autocorrelation in the autocorrelation plot, we increase the total number of samples by generating $h \times L$ samples after convergence and then saving hth sample to calculate the posterior descriptive statistics. In this way, the total number of samples after thinning and convergence is L with the hth thinning.

Bayesian Model Selection, Evaluation, and Testing

Model selection. One model was selected among four models regarding dimensionality (i.e., one random person effect across categories vs. category-specific random person effects) and random lag slopes (i.e., $\boldsymbol{\lambda}_{R.j}$ for the model at $t = 2, \ldots, T$). The linear predictors in the baseline model include one random person effect (across categories) and fixed lag effects in the linear predictors for $t > 1$, as follows:

$$\zeta_{Btji2} = (\gamma_{12} + \theta_{1j} + \beta_{1i2})I[t = 1] + (\gamma_2 + C'_{(t-1)j2}\lambda_{22} + C'_{(t-1)j3}\lambda_{32} + \theta_j + \beta_{i2})I[t > 1], \quad (11)$$

$$\zeta_{Btji3} = (\gamma_{13} + \theta_{1j} + \beta_{1i3})I[t = 1] + (\gamma_3 + C'_{(t-1)j2}\lambda_{23} + C'_{(t-1)j3}\lambda_{33} + \theta_j + \beta_{i3})I[t > 1], \quad (12)$$

and $\zeta_{Btji1} = 0$. Equations 4 and 5 are for the model that has category-specific random person effects and random lag slope effects. For variance of a random person effect at $t = 1$ and $t > 1$ in Models 1 and 3, an inverse-gamma hyper-prior distribution was used with the parameters 1000 ($dgamma(0.001, 0.001)$ in `OpenBUGS`).

==========================

Insert Table 2 about here

==========================

In this study, the four competing models were compared using a deviance information criterion (DIC; Spiegelhalter, Best, Carlin, & van der Linde, 2002). Spiegelhalter et al. (2002) specify the DIC as the loss in assigning the model predictive density $f(\mathbf{y}^{new}|\widehat{\Theta})$ to new or out-of-sample data $\mathbf{y}^{new}$ as $-2logf(\mathbf{y}^{new}|\widehat{\Theta})$. They use the expectation of $-2logf(\mathbf{y}^{new}|\widehat{\Theta})$ as the sum of the deviance for the observed data $\mathbf{y}$ $(-2logf(\mathbf{y}|\widehat{\Theta}))$ and the *optimism* because of having the same data to estimate Θ and the loss:

$$E[-2logf(\mathbf{y}^{new}|\widehat{\Theta})] = -2logf(\mathbf{y}|\widehat{\Theta}) + optimism. \tag{13}$$

The *optimism* is defined as follows for a model m (Spiegelhalter et al., 2002):

$$optimism \simeq 2p_m = 2\{E[-2logf(\mathbf{y}|\Theta)] + 2logf(\mathbf{y}|\widehat{\Theta})\}, \tag{14}$$

where $E[-2logf(\mathbf{y}|\Theta)]$ is the posterior expectation of the deviance and $2logf(\mathbf{y}|\widehat{\Theta})$ is the deviance at the posterior mean. Adding Equation 14 to Equation 13 results in

$$DIC = -2logf(\mathbf{y}|\widehat{\Theta}) + 2p_m = E[-2logf(\mathbf{y}|\Theta)] + p_m. \tag{15}$$

Because we consider random effects as a set of the parameters ($\Theta = [\gamma_{1k}, \gamma_k, \lambda_{2k}, \lambda_{3k}, \boldsymbol{\theta}_1, \boldsymbol{\beta}_1, \boldsymbol{\theta}, \boldsymbol{\beta}, \Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4]$) in a hierarchical Bayesian model, the conditional log-likelihood for $-2logf(\mathbf{y}|\widehat{\Theta})$ in DIC was calculated by conditioning on random person effects for persons and random item effects for items in the person-by-item data. A smaller DIC represents a better fit of the model, and a difference of less than 5 or 10 units between models does not provide sufficient evidence for favoring one model over another (Spiegelhalter, Thomas, Best, & Lunn, 2003).

Model evaluation. To assess model-data fit for a selected model, posterior predictive model checking (PPMC; Gelman, Meng, & Stern, 1996) was used to examine whether the selected model described the observed data adequately to draw inferences about the parameters. The idea of the PPMC is to generate replicated data from the posterior predictive distribution for the fitted model and to compare the replicated data to the observed datasets. The selected model fits well when the replicated data are similar to the observed data.

We use a discrepancy measure based on the χ^2-statistic, given by

$$D_{\chi^2}(y_{tji}|P_{tji}) = \frac{(y_{tji} - E[y_{tji}])^2}{E[y_{tji}]}, \tag{16}$$

where $E[y_{tji}]$ is the expected category score, calculated as $y_{tji}P_{tji}$. For the l-th posterior draw, the observed and replicated discrepancies are calculated as

$$D_{\chi^2}(y_{tji}|P_{tji}^{(l)}) = \frac{(y_{tji} - E[y_{tji}]^{(l)})^2}{E[y_{tji}]^{(l)}} \tag{17}$$

and

$$D_{\chi^2}(y_{tji}^{rep(l)}|P_{tji}^{(l)}) = \frac{(y_{tji}^{rep(l)} - E[y_{tji}]^{(l)})^2}{E[y_{tji}]^{(l)}}, \tag{18}$$

where $E[y_{tji}]^{(l)} = y_{tji}P_{tji}^{(l)}$ and $y_{tji}^{rep(l)}$ is the lth replicated data, sampled from a categorical distribution, i.e., $y_{tji}^{rep(l)} \sim categorical(P_{tji}^{(l)})$. Posterior predictive p-values (ppp-values; Meng, 1994) were calculated as

$$ppp = L^{-1} \sum_{l=1}^{L} I[D_{\chi^2}(y_{tji}^{rep(l)}|P_{tji}^{(l)}) \geq D_{\chi^2}(y_{tji}|P_{tji}^{(l)})], \tag{19}$$

where l is an index for draw after burn-in ($l = 1, \ldots, L$). We consider a ppp-value smaller than .025 or larger than .975 as extreme values indicative of misfit at the 5% level.

Testing. As a posterior moment for fixed effects of interest (i.e., fixed lag effects, λ_{1k} and λ_{2k}) and variance in Σ_3, the posterior mean, the posterior median, and the 95% highest posterior density (HPD) interval were reported based on iterations after discarding a burn-in period. Significance of the effects was tested using the 95% HPD interval. When the HPD interval did not include 0, the effects were considered significantly different from 0.

Illustration

In this section, we illustrate a Markov mixed-effect multinomial logistic regression model using an empirical data set. The data set comes from a study previously published as Jacobs, Cho, and Watson (2019). The primary objective of analysis was to determine whether speakers are more likely to produce a DO or PO syntactic structure if it was produced on the immediately preceding trial.

Participants

600 participants were recruited from Amazon Mechanical Turk. Amazon Mechanical Turk is an online platform on which tasks are crowdsourced to individuals for pay. It is a popular participant recruitment tool among psychologists because large numbers of participants can be recruited online within a few hours (e.g., Thomas & Clifford, 2017). All participants were self-reported native speakers of English who lived in the United States and who had completed at least 1000 tasks with a 99% approval rate by experimenters. While we believe that these participants come from more diverse backgrounds than introductory psychology students at the authors' institutions (Difallah, Filatova, & Ipeirotis, 2018), we did not collect additional demographic information to confirm this. Participants received $1 in compensation for their time. We only excluded participants from the experiment who did not respond to all questions, participants who did not produce grammatical English sentences (with an exclusion criterion of 3 ungrammatical sentences from a set of 7 filler constructions), and who did not answer catch trials appropriately. 492 of the 600 remained after these exclusions.

Experimental Design and Procedure

In this study, participants described 17 images (first production phase), then rated the truthfulness of 16 images (comprehension phase), and then described an additional 17 images (second production phase). In the original study, the purpose of the comprehension phase was to demonstrate that speakers could be primed by the productions of others, which was meant to replicate previous findings (Bock, 1986) and validate the task. In the comprehension phase, participants were exposed to 6 images depicting ditransitive events (e.g., "Jenny handed David a drink."), which were described using DO or PO structures. In the control comprehension condition, participants saw a set of 6 images that differed from those presented in the production phase. These images depicted simple events that were described using a mix of active and passive voice (e.g., "A bee stung the man.").

Participants were not given specific instructions about how to describe the images. Because there were two production phases, we counterbalanced the production phases across participants. In the production phases of the experiment, there were 10 filler images and 7 critical images. The 10 filler images depicted events that were unlikely to elicit DO or PO ditransitives, instead resulting in simple intransitive descriptions (e.g., "The cats are sleeping.") or transitive descriptions ("The woman is eating an apple."). The 10 filler images were included to prevent participants from identifying the goals of the experiment and from noticing the ditransitive syntactic structures. Fillers therefore are not included in the analysis. One potential consequence of including fillers in the design, however, is that greater amounts of time pass between trials depicting DO/PO events, and so fillers may weaken any apparent self-priming. However, previous work in comprehension priming has shown that syntactic priming occurs even when prime and target are separated by intervening grammatically unrelated trials (e.g., Bock & Griffin, 2000 among many others).

The 7 critical images depicted events that could potentially be described using either a PO or DO syntactic structure (e.g., "A man hands a bouquet of flowers to a couple" or "A man hands a couple a bouquet of flowers"). Because of the counterbalancing, this means that a total of 14 images (items) are included for analysis, with roughly half of participants seeing the first set of 7 images, and the other half of the participants seeing the second set of 7 images. All of the images appeared in succession on a single page of a Qualtrics survey. We provide a schematic of the trial structure with representative images in Figure 2. To code the data, the third author in the current study manually annotated the construction used by the participants only on the seven critical trials per block, i.e., DO, PO, or Other. The coder was blind to the previous trial and the trial number.

==========================

Insert Figure 2 about here

==========================

Descriptive Analysis

In the data for analysis, there are 492 participants and 14 items, and each participant has 7 critical trials per block (i.e., the design is balanced). Intervals between trials are equal and there are no missing responses. Figure 3 shows a mean trend (across persons) over trials in empirical logit for Other vs. DO and Other vs. PO. No evident trend (e.g., increasing or decreasing patterns) in the empirical logit was observed. Because our interest is in modeling lag effects over 7 trials, transition percentage from a trial $t-1$ to a trial t was calculated. As shown in Table 3, repetition was the most likely outcome of all three types of transitions: speakers tended to repeat Other structures 45.55% of the time, DO structures 41.12% of the time, and PO structures 47.79% of the time. Based on this evidence, we consider dependencies between adjacent trials in a Markov mixed-effect multinomial logistic regression model.

==

Insert Table 3 and Figure 3 about here

==

Analysis of a Markov Mixed-Effect Multinomial Logistic Regression Model

For all four models summarized in Table 2, a conservative burn-in of 6,000 iterations was used followed by 8,000 post–burn-in iterations for parameters, based on the PSRF and autocorrelation plots. Thinning was set at 10 based on the autocorrelation plots, indicating that 80,000 iterations were required after burn-in to obtain the 8,000 iterations. Table 2 shows the DIC values among the four models. DIC was the smallest for Model 4 among the four models, indicating that there were different response tendencies between Other vs. DO and Other vs. PO and there were individual differences in lag effects. Results of fixed effects and variance-covariance of random effects of Model 4 were comparable between the data-based prior and inverse-Wishart distribution, which suggests that results from Bayesian analysis may not be affected by the prior distributions we selected. For fixed lag effects in Model 4, as an example of comparability, the posterior median for λ_{22} was 0.464 (HPD=[0.056,0.858]) in the data-based prior and it was 0.403 (HPD=[0.043,0.833]) in the inverse-Wishart distribution; the posterior median for λ_{33} was 0.392 (HPD=[0.040,0.751]) in the data-based prior and it was 0.397 (HPD=[0.048,0.716]) in the inverse-Wishart distribution.

Posterior distributions were symmetric and unimodal for all parameters of Model 4. In Table 4, the posterior mean, median, and 95% HPD are reported for fixed effects, which are the primary effects of interest. Due to space limitations, posterior median for population parameters of random effects is presented. Based on the *ppp*-values, there were 8% of total observations (3,444) having *ppp*-values $> .975$ for Model 4. Figure 4 presented observed category scores and predicted category scores (calculated as $\sum_{k=1}^{K}[\tilde{P}_{tji}^{[y_{tji}=k]}(y_{tji}=k)]$) for selected participants. The observed category scores were close to the predicted category scores for most participants. These results suggest that there is evidence that the model describes the data relatively well.

==

Insert Table 4 and Figure 4 about here

==

Results of a Markov Mixed-Effect Multinomial Logistic Regression Model

Our primary interest is in interpreting the two fixed lag effects (λ_{22} and λ_{33}) and variances of these two effects ($Var(\lambda_{j22})$ and $Var(\lambda_{j33})$) for the DO-DO and PO-PO reuse. The fixed lag effect for Other vs. DO is significant for the Other vs. DO outcome comparison (median=0.403, HPD=[0.043,0.833]), controlling for the other effects in the model. The fixed lag effect in the Other vs. PO comparison is also significant for the Other vs. PO outcome comparison (median=0.397, HPD=[0.048,0.716]), controlling for the other effects in the model. Variance estimates of the two lag effects (posterior median) were 0.296 (HPD=[0.141,0.667]) and 0.607 (HPD=[0.244,1.276]), which suggests that there are some nonignorable individual differences in the lag effects. Figure 5 shows the scatter plot of the lag effects $\boldsymbol{\lambda}_j = [\lambda_{j22}, \lambda_{j32}, \lambda_{j23}, \lambda_{j33}]$ (posterior median) for individual persons, showing that there is non-ignorable interpersonal variability in the lag effects. The λ_{j22} for DO-DO ranged from 0.107 to 0.869, which suggests that all individuals were more likely to produce a DO than Other after having just produced a DO. The λ_{j33} for PO-PO ranged from -0.354 to 1.225 and 18.09% of individuals had negative λ_{j33}. The negative λ_{j33} indicates that the odds of producing a PO compared to the Other after having just produced a PO are lower than having just produced an Other. However, a majority of subjects (81.91%) were more likely to produce a PO than Other after having just produced a PO, compared to having just produced an Other.

Simulation Study

Simulation Design and Analysis

A simulation study was designed to answer the following two questions: (a) can the parameters of the selected model (Model 4 in Table 2) and precision (i.e., posterior standard deviation) be recovered at a satisfactory level in the same condition as the empirical study (7 trials, 14 items, and 492 persons)?, and (b) to what extent can the accuracy of parameter estimates and precision (especially, the accuracy of the focal lag effects (λ_{22}, λ_{33}, Var(C_1 for Other vs. DO), and Var(C_2 for Other vs. PO)) be improved in a larger number of trials, items, and persons than in the condition of the empirical study? For the second question, the number of trials, persons, and items was chosen with the experimental design of the self-priming effects in mind, that is, the number of trials is equal to half of the number of items. In addition, there are practical constraints of data collection using Amazon Mechanical Turk. Mechanical Turk participants are essentially paid volunteers who complete studies online. Most Mechanical Turk participants are looking to maximize monetary payout while minimizing time spent on a task. Considering this constraint, the largest number of items and persons used for the second question were 42 items and 1,000 persons. To summarize, the following conditions were considered to answer the two questions:

- The number of trials and items, presented as (the number of trials, the number of items) in the following: (7, 14), (14, 28), and (21, 42). The level of (14, 28) was considered to be as a medium level because 42 items are the largest number of items we can consider in practice.

- The number of persons: 492 persons, 750 persons, and 1,000 persons. The level of 750 persons was chosen as a medium level when 1,000 persons are the largest number of persons we can consider in practice.

The listed two conditions above are fully crossed, resulting in 9 conditions. Five hundred replications for each condition were considered.

For data generation in the simulation study, estimates (posterior medians) of Model 4 in Table 4 were considered true parameters and the random item assignment to persons and trials was used as in the empirical study. Mahowald et al. (2016) conducted a meta analysis of 73 peer-reviewed journal papers on syntactic priming and reported that the syntactic priming effect in the empirical study is small. Thus, we did not consider stronger magnitudes of the first-order lag parameters with which we interpreted the syntactic self-priming effects in the simulation study.

The same prior and hyper-prior distributions used in the empirical study were used in the simulation study using `OpenBUGS` for Bayesian analysis. No convergence problems were encountered in any replication for all conditions. In order to check convergence, 10% of the 500 replications were used. Based on the PSRF and autocorrelation plots, burn-in of 6,000-8,000 iterations varied across conditions was used, followed by 8,000 post–burn-in iterations ($L = 8,000$) with a thinning factor (h) of 5-10 varied across conditions, meaning that $8,000 \times h$ iterations were required after burn-in to obtain the 8,000 iterations. For all parameter estimates, Monte Carlo standard error (MCSE) was smaller than 1% of the standard deviations of the estimates (SD).

Posterior mean was used to calculate two accuracy measures, bias and root mean square error (RMSE). In addition, the relative percentage bias was also used to evaluate whether bias for parameters of interest (λ_{22}, λ_{33}, Var(C_1 for Other vs. DO), and Var(C_2 for Other vs. PO)) is acceptable with an empirical cutoff. We consider the relative percentage bias up to |10|% as an acceptable level (L. K. Muthén & Muthén, 2002). To evaluate whether the estimated standard deviation is approximately correct, the mean standard error estimates (M(SE)) across 500 replications was compared with SD.

Simulation Results

Results of Question (a). Results of Question (a) are presented in Table 5. The following overall patterns were evident. First, bias and RMSE for fixed and population parameters of random effects were smaller for the model at $t > 1$ than for the model at $t = 1$, which is not surprising because parameters at $t = 1$ were estimated based on data from the first trial only. Second, bias and RMSE were smaller for fixed effects than for the population parameters of random effects. Third, parameters of interest (λ_{22}, λ_{33}, Var(C_1 for Other vs. DO), and Var(C_2 for Other vs. PO)) were recovered relatively well, compared to the other parameters in the model. Fourth, M(SE) across 500 replications approached SD, suggesting that the estimated posterior standard deviations are approximately correct (range of $SD/M(SE) = [0.945, 1.050]$ for the model at $t = 1$; range of $SD/M(SE) = [0.933, 1.084]$ for the model at $t > 1$). To conclude, some parameter estimates have a large bias and RMSE, especially for the model at $t = 1$ given the condition of 7 trials, 492 persons, and 14 items. However, parameter recovery of the lag effects we focus on was relatively acceptable: relative percentage bias for λ_{22}, λ_{33}, Var(C_1 for Other vs. DO), and Var(C_2 for Other vs. PO) were −1.015%, −10.579%, 10.811%, and −8.567%, respectively. In addition, the precision for these parameters was satisfactory: $SD/M(SE)$ for λ_{22}, λ_{33}, Var(C_1 for Other vs. DO), and Var(C_2 for Other vs. PO) were 0.957, 1.006, 1.018, and 0.975, respectively.

=========================

Insert Table 5 about here

=========================

Results of Question (b). The results of Question (b) are shown in Table 6 and Table 7. Average bias (top in Table 6) and RMSE (bottom in Table 6) are reported by levels of the two conditions to patterns in bias and RMSE by increasing the number of trials, items, and persons. The following patterns were observed. First, overall, bias and RMSE for fixed and population parameters of random effects were smaller for the model at $t > 1$ than for the model at $t = 1$ in all conditions, as in the results of Question (a). Second, as expected, bias and RMSE of all parameter estimates were smaller, with an increasing number of trials, items, and persons. Third, as presented in Table 7 (top), the relative percentage bias was less than |10.214|% for focal lag effects (λ_{22}, λ_{33}, Var(C_1 for Other vs. DO), and Var(C_2 for Other vs. PO)) and it decreased with a larger number of trials, items, and persons. On average (across four focal parameter estimates), the absolute value of the relative percentage bias decreased from 7.364% to 2.194%, with an increasing number of trials and items from (7,14) to (14,28). In addition, it decreased from 5.622% to 1.784%, with an increasing number of persons from 492 to 1000. Furthermore, $SD/M(SE)$ is close to 1 as shown in Table 7 [bottom]). To summarize, accuracy and precision for the focal lag effects were acceptable in all levels of the conditions we considered, and improved with a larger number of trials, items, and persons.

==

Insert Table 6 and Table 7 about here

==

Summary and Discussion

Although syntactic priming effects have been investigated for several decades, appropriate statistical modeling has not been applied to understanding self-priming in the literature. This paper presents a Markov mixed-effect multinominal logistic regression model in which the syntactic self-priming effect was modelled with the covariates of own-category lags (i.e., DO-DO, PO-PO) and cross-category lags (i.e., DO-PO, PO-DO). The model specification in the present study has the potential for studying the level (average) and individual differences (participant-specific deviations) in syntactic self-priming effects when dependency due to items should be modelled in the person-by-item data. This potential was illustrated using a data set from an experimental study in psycholinguistics. We find evidence of self-priming, which is consistent with the predictions of activation-based theories. These theories claim that previous use of syntactic structures prime their activation, independent of who uttered it, which means that speakers will be more likely to produce a primed structure both when the prime is produced by the speaker or by someone else. Self-priming is less consistent with the expectation-based theories, where the likelihood of priming increases after encounters with unexpected linguistic structures. By definition, a speaker's own productions cannot be unexpected. Thus, the expectation-based theory must be revised to accommodate these findings. More generally, a Markov mixed-effect multinominal logistic regression model has allowed us to address the predictions of two important theories in psycholinguistics.

This study focused on both the improvements to existing modeling methodology in order to deepen understanding of syntactic self-priming and on its application to the data. Thus, several methodological limitations remain in the current study, presented in order below. The simulation conditions we considered are limited. Our initial investigation of the Bayesian analysis in `OpenBUGS` showed that accuracy and precision of estimates for focal lag effects (λ_{22}, λ_{33}, Var(C_1 for Other vs. DO), and Var(C_2 for Other vs. PO)) can be acceptable in the same conditions as the empirical study and the other conditions of other possible experimental designs found in practice. However, a large-scale simulation study is required to generalize our simulation results to other conditions that vary in the number of trials, persons, and items, and magnitudes of the first-order lag parameters.

A model was selected based on the DIC among four models with which we investigated whether or not there are different response tendencies between Other vs. DO and Other vs. PO and whether or not there are individual differences in lag effects. We used the conditional log-likelihood instead of the marginal log-likelihood for the DIC calculation because random effects are a subset of parameters of interest (e.g., $\boldsymbol{\lambda}_j$), which means that the new data $\mathbf{y}^{new}$ in hypothetical replications come from the same clusters as the observed data. It has been noted that the $2p_m$ may not be a good approximation of the *optimism* in Equation 14 in the use of the conditional log-likelihood because the dimensionality of the parameters increases with the sample size (Merkle, Furr, & Rebe-Hesketh, 2019; Plummer, 2008). Further research is required to investigate the performance of the DIC with the conditional log-likelihood, in selecting a 'true' set of the random effects in various simulation conditions regarding the number of persons, items, trials, and magnitude of variance and covariance for the random effects.

We considered first-order lag effects for our motivating example. The model adequacy analysis using the PPMC provides evidence that the first-order fixed and random lag effects described the observed data adequately without including the higher-order lag effects and trial-varying lag effects. Considering higher-order lag effects in the Markov mixed-effect multinomial logistic models is straightforward as can be seen from our illustration of generating the first-order lag effects for nominal responses (see Table 1). However, having higher-order random lag effects increases the number of random effects and thus is computationally expensive. Another interesting extension is to consider trial-varying lag effects in the model to investigate trial (or time)-varying lag effects (e.g., Bringmann et al., 2017). As found in the empirical study of Pettitt et al. (2006), having random lag effects and trial-varying lag effects resulted in convergence problems in estimation in our application, which may indicate over-fit serial dependence in the data. However, we recommend considering possible trial-varying lag effects as a candidate model, especially when the number of trials is large (e.g., > 30 as in time-series data).

Although there are several methodological limitations that still need to be investigated in further studies, it is our hope that this paper provides a new model-based perspective to investigate priming effects that have been the subject of investigation for several decades. To our knowledge, this paper is the first to detect effects of syntactic self-priming using lag covariates in the regression-type model, by accounting for multiple sources of dependency (e.g., persons and items) and variability (e.g., individual differences in the own-category and crossed-category lag effects) in the literature. We anticipate that the model we present in this paper will enhance data analysis practices for investigating syntactic priming effects. Methodologically, the model specification for nominal item response data in the present study can be applicable to that of ordinal item response data. Although interpretations of lag parameters are different in a baseline logit for the nominal response data compared to an adjacent-category or cumulative logit for the ordinal item response data (see Agresti [2013] for different logit interpretations), we hope that this paper can serve as an example of the model specification to model fixed and random own-category lags and cross-category lag effects for ordinal person-by-item data.

Table 1: Coding for Covariates

response	y_{tji}	C_{tj2}	C_{tj3}
Other	1	0	0
DO	2	1	0
PO	3	0	1

trial t	person j	item i	y_{tji}	C_{tj2}	C_{tj3}	$C_{(t-1)j2}$	$C_{(t-1)j3}$	$I[t=1]$	$I[t>1]$
1	1	13	2	1	0	.	.	1	0
2	1	10	3	0	1	1	0	0	1
3	1	11	3	0	1	0	1	0	1
4	1	12	1	0	0	0	1	0	1
5	1	14	1	0	0	0	0	0	1
6	1	5	1	0	0	0	0	0	1
7	1	2	2	1	0	0	0	0	1
1	2	2	3	0	1	.	.	1	0
2	2	14	1	0	0	0	1	0	1
3	2	13	1	0	0	0	0	0	1
4	2	12	1	0	0	0	0	0	1
5	2	10	1	0	0	0	0	0	1
6	2	5	1	0	0	0	0	0	1
7	2	11	3	0	1	0	0	0	1
1	3	4	3	0	1	.	.	1	0
2	3	9	3	0	1	0	1	0	1
3	3	8	3	0	1	0	1	0	1
4	3	1	2	1	0	0	1	0	1
5	3	7	3	0	1	1	0	0	1
6	3	6	1	0	0	0	1	0	1
7	3	3	3	0	1	0	0	0	1

Note. "." indicates a missing value.

Table 2: Empirical Study: Model Selection using DIC

Model (m)	Category-Specific?	Slope?	$D(\widehat{\Theta}_m, m)$	p_m	DIC
Model 1 ($m1$)	No	No	6280.18	309.53	6589.71(3)
Model 2 ($m2$)	Yes	No	5657.88	763.31	6421.19(2)
Model 3 ($m3$)	No	Yes	6128.58	472.21	6600.79(4)
Model 4 ($m4$)	Yes	Yes	5616.08	793.11	6409.19(1)

Note. $D(\widehat{\Theta}_m, m)$ is deviance at posterior means ; p_m is the effective number of parameters for model m; numbers in parenthesis for DIC indicate model ranks from the smallest DIC to the largest DIC.

Table 3: Empirical Study: Descriptive Transition Percentage from Trial $t-1$ to Trial t

		y_{tji}			
		1	2	3	Total
$y_{(t-1)ji}$	1	**45.55**	22.90	31.54	100.00
	2	35.77	**41.12**	23.11	100.00
	3	33.00	19.22	**47.79**	100.00

Note. $y_{tji} = 1$ for Other; $y_{tji} = 2$ for DO; $y_{tji} = 3$ for PO.

Table 4: Empirical Study: Results of Fixed Effects (Top) and Random Effects (Bottom)

	Other vs. DO Outcome			Other vs. PO Outcome		
	Mean	Median	HPD	Mean	Median	HPD
$t = 1$						
Grand Mean[γ_{1k}]	**-0.728**	**-0.716**	[-1.461,-0.045]	-0.365	-0.358	[-1.048,0.240]
$t > 1$						
Mean for Other[γ_k]	-0.599	-0.607	[-1.205,0.134]	-0.155	-0.149	[-0.619,0.279]
$C_{(t-1)j1}$ for Other vs. DO[λ_{k2}]	**0.394**	**0.403**	[0.043,0.833]	-0.333	-0.334	[-0.738,0.068]
$C_{(t-1)j2}$ for Other vs. PO[λ_{k3}]	-0.027	-0.021	[-0.456,0.390]	**0.395**	**0.397**	[0.048,0.716]

	Median Variance	Covariance				
$t = 1$						
Persons [Σ_1]						
Other vs. DO	0.415					
Other vs. PO	0.367	-0.017				
Items [Σ_2]						
Other vs. DO	1.121					
Other vs. PO	1.098	0.807				
$t > 1$						
Persons [Σ_3]						
Other vs. DO	1.009					
Other vs. PO	1.778	0.358				
$C_{(t-1)j1}$ for Other vs. DO	0.296	-0.008	0.030			
$C_{(t-1)j1}$ for Other vs. PO	0.399	0.288	-0.069	0.006		
$C_{(t-1)j2}$ for Other vs. DO	0.373	0.254	-0.103	0.002	0.123	
$C_{(t-1)j2}$ for Other vs. PO	0.607	-0.111	-0.797	-0.027	0.078	0.107
Items [Σ_4]						
Other vs. DO	1.211					
Other vs. PO	0.716	0.844				

Note. Significance in bold for fixed effects based on 95% HPD.

Table 5: Simulation Results of Question (a): Accuracy of Parameter Estimates and Precision

	Parameters	Bias	RMSE	SD	M(SE)
Fixed					
$t = 1$					
	γ_{12}	-0.234	0.459	0.395	0.412
	γ_{13}	-0.090	0.401	0.391	0.369
$t > 1$					
	γ_2	-0.043	0.305	0.302	0.291
	γ_3	0.033	0.251	0.249	0.239
	$\underline{\lambda_{22}}$	-0.004	0.157	0.157	0.164
	λ_{32}	-0.065	0.189	0.177	0.169
	λ_{23}	0.022	0.162	0.160	0.172
	$\underline{\lambda_{33}}$	-0.042	0.145	0.139	0.138
Random					
$t = 1$					
Persons $[\Sigma_1]$	Var(Other vs. DO)	0.211	0.405	0.346	0.366
	Var(Other vs. PO)	0.259	0.488	0.414	0.418
	Covariance	-0.286	0.502	0.413	0.421
Items $[\Sigma_2]$	Var(Other vs. DO)	0.147	0.704	0.688	0.717
	Var(Other vs. PO)	0.224	0.540	0.491	0.500
	Covariance	-0.005	0.550	0.550	0.576
$t > 1$					
Persons $[\Sigma_3]$	Var(Other vs. DO)	0.053	0.291	0.286	0.270
	Var(Other vs. PO)	-0.165	0.320	0.274	0.253
	<u>Var(C_1 for Other vs. DO)</u>	0.032	0.180	0.177	0.174
	Var(C_1 for Other vs. PO)	0.073	0.158	0.140	0.149
	Var(C_2 for Other vs. DO)	0.151	0.262	0.214	0.213
	<u>Var(C_2 for Other vs. PO)</u>	-0.052	0.236	0.230	0.236
	Covariance	0.017	0.207	0.206	0.199
Items $[\Sigma_4]$	Var(Other vs. DO)	0.030	0.486	0.485	0.500
	Var(Other vs. PO)	0.101	0.313	0.296	0.300
	Covariance	-0.219	0.380	0.311	0.309

Note. For covariance in Σ_3, results were averaged across 15 covariance terms; Parameters of interest were underlined.

Table 6: Simulation Results of Question (b): Bias (top) and RMSE (bottom) of All Parameter Estimates

	Parameters	Conditions					
		(Trials, Items)			Persons		
		(7,14)	(14,28)	(21,42)	492	750	1000
Fixed							
$t = 1$							
	γ_{12}	-0.187	-0.173	-0.110	-0.182	-0.182	-0.112
	γ_{13}	-0.066	-0.057	-0.048	-0.147	-0.092	-0.088
$t > 1$							
	γ_2	-0.039	0.031	0.029	-0.035	0.005	0.003
	γ_3	-0.025	0.015	-0.068	-0.014	-0.011	-0.003
	$\underline{\lambda_{22}}$	0.003	0.003	0.001	0.003	0.002	0.000
	$\underline{\lambda_{32}}$	0.022	0.013	0.005	0.039	0.030	0.001
	λ_{23}	0.019	0.016	0.010	0.020	0.018	0.011
	$\underline{\lambda_{33}}$	-0.033	0.028	0.009	0.035	0.034	-0.002
Random							
$t = 1$							
Persons $[\Sigma_1]$	Var(Other vs. DO)	0.210	0.193	0.090	0.214	0.199	0.095
	Var(Other vs. PO)	0.253	0.247	0.131	0.244	0.201	0.098
	Covariance	-0.285	-0.277	-0.069	-0.276	-0.275	-0.089
Items $[\Sigma_2]$	Var(Other vs. DO)	0.147	0.107	0.059	0.145	0.109	0.044
	Var(Other vs. PO)	0.220	0.196	0.061	0.268	0.266	0.059
	Covariance	-0.005	-0.004	0.001	-0.004	-0.004	0.002
$t > 1$							
Persons $[\Sigma_3]$	Var(Other vs. DO)	-0.072	0.021	-0.015	-0.040	-0.020	-0.001
	Var(Other vs. PO)	-0.172	-0.021	-0.008	-0.186	-0.177	-0.094
	Var(C_1 for Other vs. DO)	0.030	0.015	-0.003	-0.012	-0.011	-0.005
	Var(C_1 for Other vs. PO)	0.027	-0.003	0.030	-0.023	-0.005	-0.008
	Var(C_2 for Other vs. DO)	0.132	-0.033	-0.014	0.034	0.026	0.024
	Var(C_2 for Other vs. PO)	-0.062	-0.043	-0.032	-0.054	-0.040	-0.030
	Covariance	0.019	0.010	0.005	0.015	0.009	0.007
Items $[\Sigma_4]$	Var(Other vs. DO)	0.021	0.015	0.009	0.053	0.021	-0.056
	Var(Other vs. PO)	0.106	0.105	0.083	0.121	0.110	0.057
	Covariance	-0.203	-0.153	-0.071	-0.246	-0.134	-0.094

	Parameters	Conditions					
		(Trials, Items)			Persons		
		(7,14)	(14,28)	(21,42)	492	750	1000
Fixed							
$t = 1$							
	γ_{12}	0.423	0.415	0.299	0.445	0.442	0.297
	γ_{13}	0.345	0.301	0.293	0.391	0.311	0.237
$t > 1$							
	γ_2	0.301	0.194	0.179	0.188	0.153	0.142
	γ_3	0.218	0.176	0.122	0.221	0.209	0.136
	$\underline{\lambda_{22}}$	0.111	0.087	0.083	0.129	0.090	0.062
	$\underline{\lambda_{32}}$	0.145	0.080	0.062	0.121	0.083	0.080
	λ_{23}	0.140	0.115	0.090	0.162	0.117	0.066
	$\underline{\lambda_{33}}$	0.128	0.092	0.062	0.107	0.102	0.074
Random							
$t = 1$							
Persons $[\Sigma_1]$	Var(Other vs. DO)	0.402	0.295	0.285	0.404	0.301	0.280
	Var(Other vs. PO)	0.411	0.310	0.207	0.410	0.309	0.226
	Covariance	0.500	0.363	0.212	0.503	0.393	0.217
Items $[\Sigma_2]$	Var(Other vs. DO)	0.698	0.496	0.232	0.665	0.499	0.221
	Var(Other vs. PO)	0.539	0.434	0.220	0.527	0.417	0.216
	Covariance	0.540	0.436	0.292	0.533	0.428	0.267
$t > 1$							
Persons $[\Sigma_3]$	Var(Other vs. DO)	0.216	0.129	0.123	0.197	0.151	0.119
	Var(Other vs. PO)	0.349	0.216	0.062	0.236	0.219	0.172
	Var(C_1 for Other vs. DO)	0.101	0.090	0.086	0.132	0.081	0.073
	Var(C_1 for Other vs. PO)	0.148	0.099	0.085	0.126	0.110	0.096
	Var(C_2 for Other vs. DO)	0.195	0.075	0.067	0.129	0.116	0.092
	Var(C_2 for Other vs. PO)	0.182	0.178	0.128	0.226	0.140	0.122
	Covariance	0.211	0.146	0.101	0.201	0.195	0.114
Items $[\Sigma_4]$	Var(Other vs. DO)	0.408	0.248	0.137	0.471	0.231	0.104
	Var(Other vs. PO)	0.310	0.150	0.116	0.325	0.202	0.111
	Covariance	0.370	0.286	0.184	0.398	0.323	0.174

Note. For covariance in Σ_3, results were averaged across 15 covariance terms; Parameters of interest were underlined.

Table 7: Simulation Results of Question (b): Relative Percentage Bias (top) and SD/M(SE) (bottom) of Focal Lag Effect Estimates

Parameters	Conditions					
	(Trials, Items)			Persons		
	(7,14)	(14,28)	(21,42)	492	750	1000
λ_{22}	0.794	0.670	0.223	0.720	0.571	0.000
λ_{33}	-8.312	7.053	2.267	8.816	8.564	-0.504
Var(C_1 for Other vs. DO)	10.135	5.068	-1.014	-4.054	-3.716	-1.689
Var(C_2 for Other vs. PO)	-10.214	-7.084	-5.272	-8.896	-6.590	-4.942

Parameters	Conditions					
	(Trials, Items)			Persons		
	(7,14)	(14,28)	(21,42)	492	750	1000
λ_{22}	0.981	0.993	1.002	0.991	1.008	1.010
λ_{33}	0.966	0.979	0.999	0.978	0.975	0.998
Var(C_1 for Other vs. DO)	0.973	0.961	0.973	0.996	0.968	0.998
Var(C_2 for Other vs. PO)	0.964	0.975	1.001	0.973	0.981	0.983

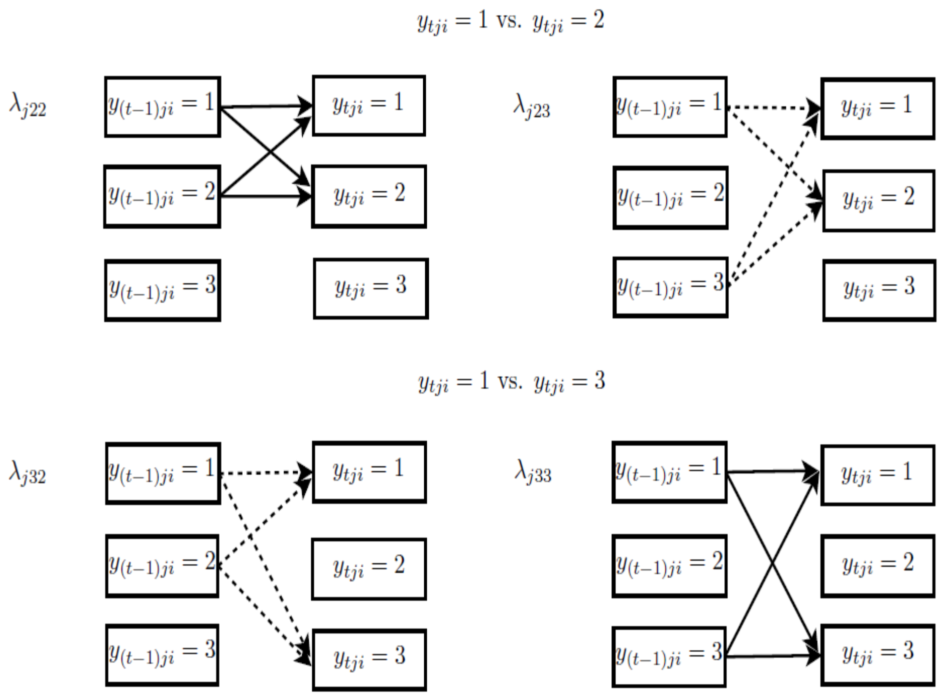

Figure 1: Graphical representation for $\boldsymbol{\lambda}_j = [\lambda_{j22}, \lambda_{j32}, \lambda_{j23}, \lambda_{j33}]$. Paths from time point $t-1$ to time point t indicate the comparison structure. λ_{j22} for Other vs. DO ($y_{tji} = 1$ vs. $y_{tji} = 2$) and λ_{j33} for Other vs. PO ($y_{tji} = 1$ vs. $y_{tji} = 3$) are for the effects of own-category lags, as presented with lines. λ_{j32} for Other vs. DO ($y_{tji} = 1$ vs. $y_{tji} = 2$) and λ_{j23} for Other vs. PO ($y_{tji} = 1$ vs. $y_{tji} = 3$)are for the effects of cross-category lags, as presented with dotted lines.

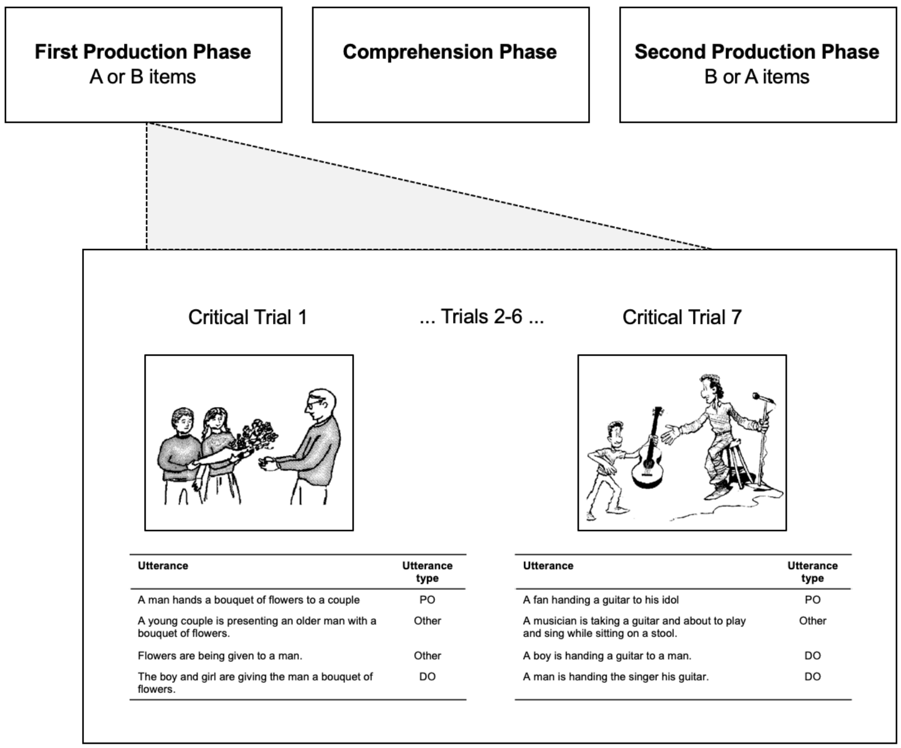

Figure 2: Representative images and trial structure from Experiment data, first block. Trials were randomly ordered. Trials were counterbalanced by participant.

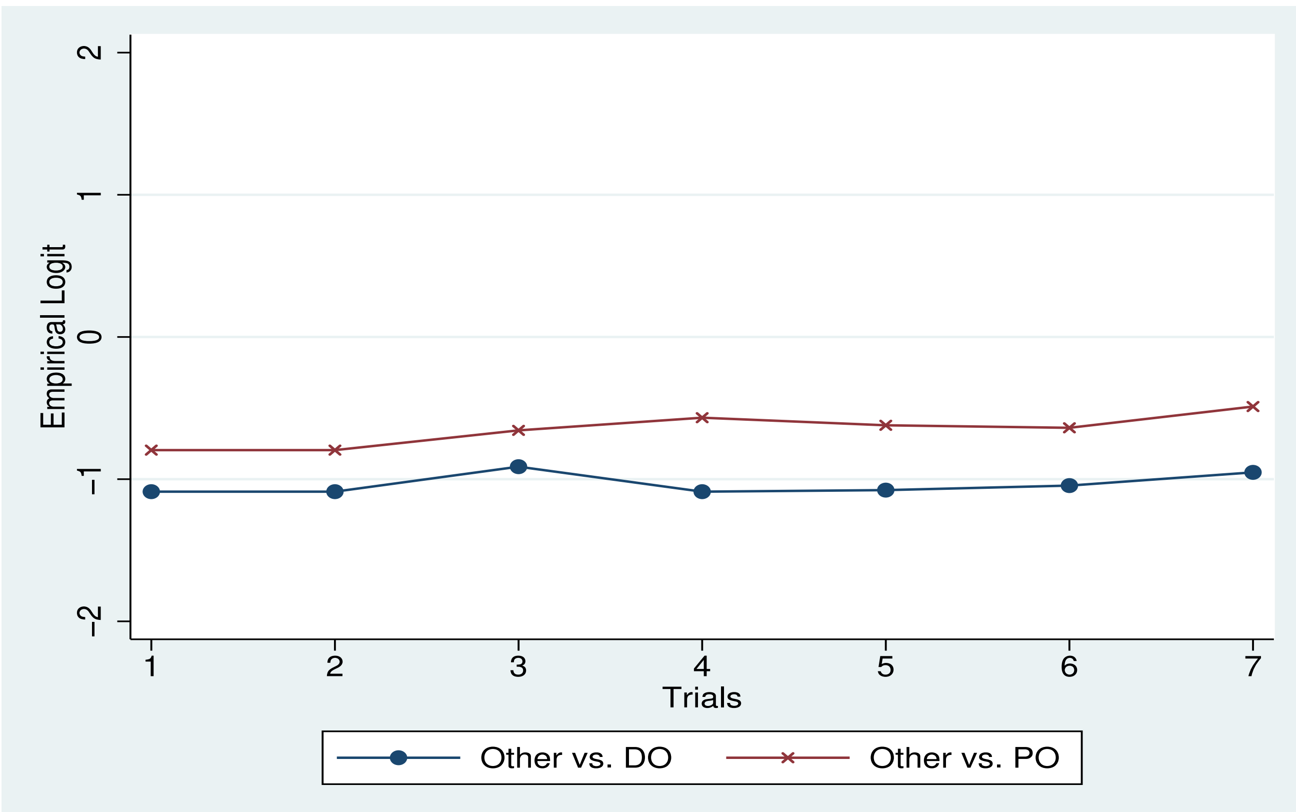

Figure 3: Mean trend over trials in empirical logit for Other vs. DO and Other vs. PO.

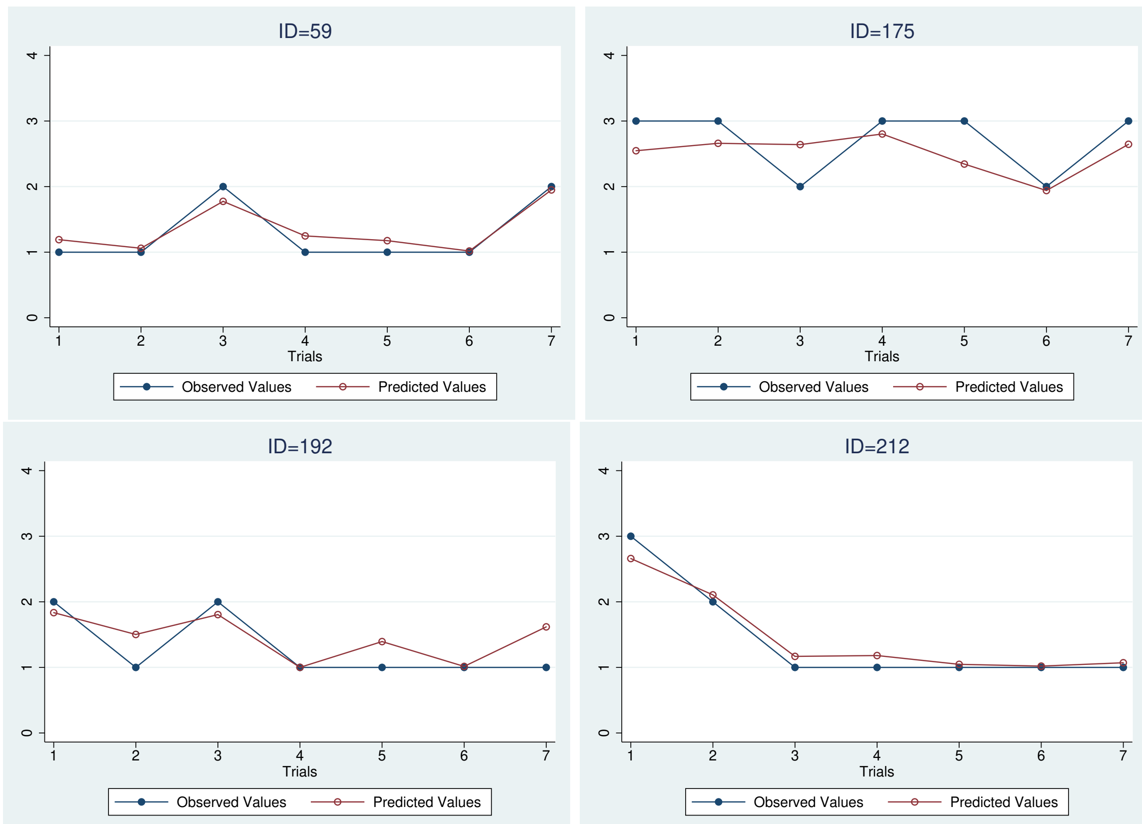

Figure 4: Empirical study: Observed vs. predicted values (calculated as $\sum_{k=1}^{K}[\tilde{P}_{tji}^{[y_{tji}=k]}(y_{tji}=k)]$) for selected participants.

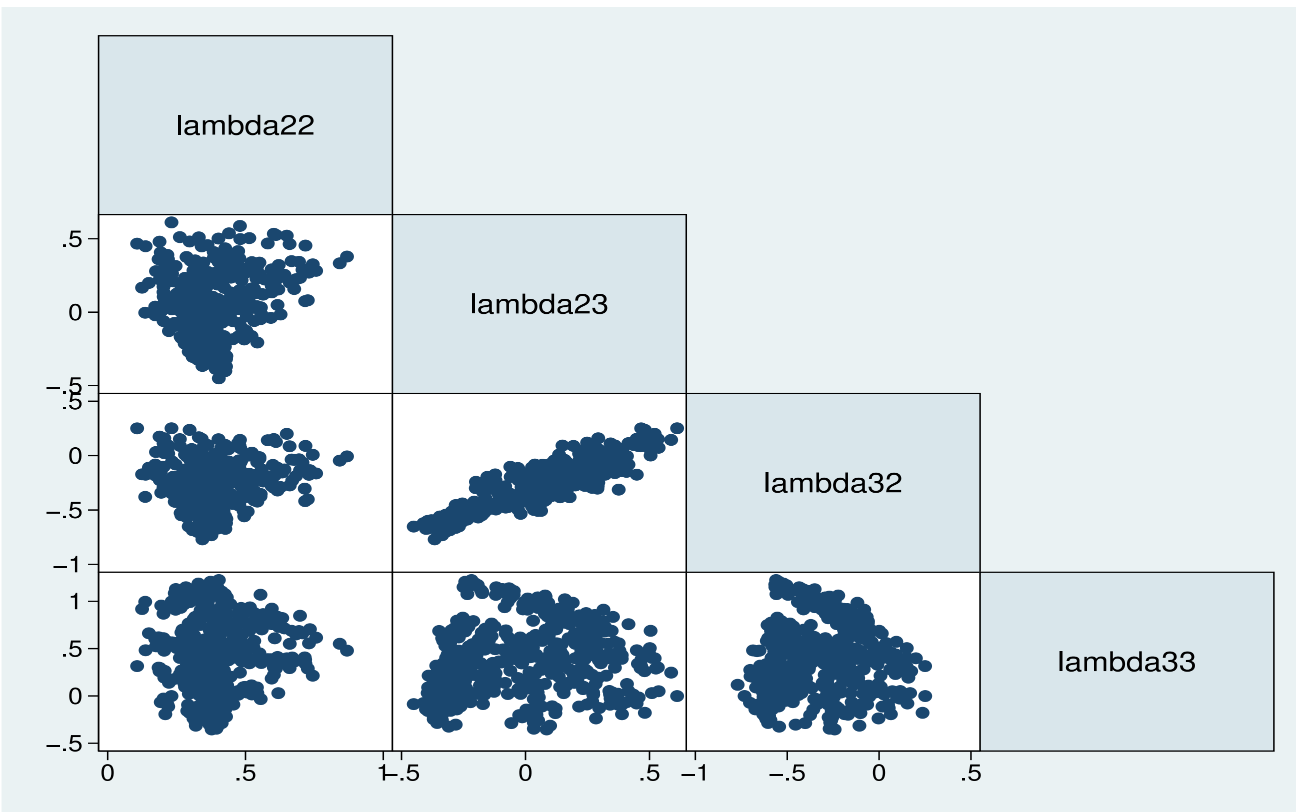

Figure 5: Empirical study: Scatter plot of the lag effects $\boldsymbol{\lambda}_j = [\lambda_{j22}, \lambda_{j32}, \lambda_{j23}, \lambda_{j33}]$ (posterior median) for the individual persons.

References

Agresti, A. (2013). *Categorical data analysis.* (3rd Ed.) New Jersey: John Wiley & Sons, Inc.

Azzalini, A. (1994). Logistic regression for autocorrelated data with application to repeated measures. *Biometrika, 81,* 767–775. doi: 10.2307/2337079

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412. doi: 10.1016/j.jml.2007.12.005

Baese-Berk, M. M. (2019). Interactions between speech perception and production during learning of novel phonemic categories. *Attention, Perception, & Psychophysics, 81,* 981–1005. doi: 10.3758/s13414-019-01725-4

Bates, D., Mächler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., . . . Green, P. (2018). *Package "lme4": Linear mixed-effects models using 'eigen' and s4.* Retrieved from https://cran.r-project.org/web/packages/lme4/lme4.pdf

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18*, 355–387. doi:10.1016/0010-0285(86)90004-6

Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: transient activation or implicit learning? *Journal of Experimental Psychology: General, 129,* 177–192. doi:10.1037/0096-3445.129.2.177

Bock, R. D. (1970). Estimating multinomial response relations. In Contributions to Statistics and Probability, Bose RC (ed.). University of North Carolina Press: Chapel Hill, NC.

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika, 37,* 29–51. doi: 10.1007/BF02291411

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1482–1493. doi: 10.1037//0278-7393.22.6.1482

Bringmann, L. F., Hamaker, E. L., Vigo, D. E., Aubert, A., Borsboom, D., & Tuerlinckx, F. (2017). Changing dynamics: Time-varying autoregressive models using generalized additive modeling. *Psychological Methods, 22*, 409–425. doi: 10.1037/met0000085

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review, 113*, 234–272. doi:10.1037/0033-295X.113.2.234

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*(4), 335–359. doi:10.1016/s0022-5371(73)80014-3

Cohen, R. G., & Rosenbaum, D. A. (2004). Where grasps are made reveals how grasps are planned: generation and recall of motor plans. *Experimental Brain Research, 157,* 486–495. doi:10.1007/s00221-004-1862-9

Cox, D. R. (1970). *The analysis of binary data.* London: Chapman and Hall.

Cox, D. R. (1981). Statistical analysis of time series: Some recent developments. *Scandinavian Journal of Statistics 8,* 93–115.

de Haan-Rietdijk, S., Gottman, J. M., Bergeman, C. S., & Hamaker, E. L. (2016). Get over it! A multilevel threshold autoregressive model for state-dependent affect regulation. *Psychometrika, 81*, 217–241. doi: 10.1007/s11336-014-9417-x

de Haan-Rietdijk, S, Kuppens, P, Bergeman, C. S., Sheeber, L. B., Allen, N. B., & Hamaker, E. L. (2017). On the use of mixed Markov models for intensive longitudinal data. *Multivariate Behavioral Research, 52,* 747–767. doi: 10.1080/00273171.2017.1370364

Difallah, D. E., Filatova, E., & Ipeirotis, P. G. (2018). Demographics and dynamics of Mechanical Turk workers. *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining,* 135–143. doi:10.1145/3159652.3159661

Fokianos, K, & Kedem, B. (2003). Regression theory for categorical time series. *Statistical Science, 18,* 357–376. doi: 10.1214/ss/1076102425

Gelman, A., & Hill, J. (2007). Data analysis using regression and multilevel/hierarchical models. Cambridge: Cambridge University Press.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7* 457–472. doi: 10.1214/ss/1177011136

Gelman, A., Meng, X. L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica, 6,* 733–759.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D., & Rubin, D. B. (2013). *Bayesian Data Analysis* (3rd ed.). Chapman & Hall/CRC.

Geyer, C. J. (1992). Practical Markov chain Monte Carlo. *Statistical Science, 7,* 473–483.

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Studies in emotion and social interaction. Contexts of accommodation: Developments in applied sociolinguistics* (pp. 1-68). New York, NY, US: Cambridge University Press.

Gries, S. T. (2005). Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research, 34*, 365–399. doi:10.1007/s10936-005-6139-3

Gruberg, N., Ostrand, R., Momma, S., & Ferreira, V. S. (2019). Syntactic entrainment: The repetition of syntactic structures in event descriptions. *Journal of Memory and Language, 107,* 216–232. doi:10.1016/j.jml.2019.04.005

Hadfield, J. D. (2010). MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *Journal of Statistical Software, 33,* 1–22. Retrieved from http://www.jstatsoft.org/v33/i02/ doi:10.18637/jss.v033.i02

Heckman, J. J. (1981). The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In C. F. Manski & D. MacFadden (Eds.), *Structural analysis of discrete data with econometric applications* (pp. 179–195). Cambridge: MIT Press.

Hedeker, D. (2003). A mixed-effects multinomial logistic regression model. *Statistics in Medicine, 22,* 1433–1446. doi: 10.1002/sim.1522

Hedeker, D. (2008). Multilevel models for ordinal and nominal variables. In J. de Leeuw & E. Meijer (Eds.), *A Handbook of Multilevel Analysis.* Springer, New York.

Jacobs, C. L., Cho, S.-J., & Watson, D. G. (2019). Self-priming in production: Evidence for a hybrid model of syntactic priming. *Cognitive Science, 43,* e12749. doi:10.1111/cogs.12749

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434–446. doi:10.1016/j.jml.2007.11.007

Korn, E. L., & Whittemore, A. S. (1979). Methods for analyzing panel studies of acute health effects of air pollution. *Biometrics 35,* 795–802. doi: 10.2307/2530111

Lodewyckx, T., Tuerlinckx, F., Kuppens, P., Allen, N. B., & Sheeber, L. B. (2011). A hierarchical state space approach to affective dynamics. *Journal of Mathematical Psychology, 55,* 68–83. doi:10.1016/j.jmp.2010.08.004

Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The BUGS project: Evolution, critique and future directions. *Statistics in Medicine, 28,* 3049–3067. doi: 10.1002/sim.3680

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology, 4,* 1–16. doi: 10.3389/fpsyg.2013.00226

MacDonald, M. C., & Weiss, D. J. (2017). Structural priming, action planning, and grammar. A commentary on Branigan & Pickering: An experimental approach to linguistic representation. *Brain & Behavioral Science, 40.* doi:10.1017/S0140525X17000498

Mahowald, K., James, A., Futrell, R., & Gibson, E. (2016). A meta-analysis of syntactic priming in language production. *Journal of Memory and Language, 91*, 5-27. doi:10.1016/j.jml.2016.03.009

Meng, X.-L. (1994). Posterior predictive p-values. *Annals of Statistics, 22,* 1142–1160. doi:10.1214/aos/1176325622

Merkle, E. C., Furr, D., & Rabe-Hesketh, S. (2019). Bayesian comparison of latent variable models: Conditional versus marginal likelihoods. *Psychometrika, 84,* 802–829. doi: 10.1007/s11336-019-09679-0

Muthén, L. K., & Muthén, B. O. (1998-2017). Mplus User's Guide. Eighth Edition. Los Angeles, CA: Muthén & Muthén.

Muthén, L. K. and Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling: A Multidisciplinary Journal, 9,* 599–620. doi:10.1207/S15328007SEM09048

Pettitt, A., Tran T., Haynes M., & Hay J. (2006). A Bayesian hierarchical model for categorical longitudinal data from a social survey of immigrants. *Journal of Royal Statistical Society-Series A., 169*, 97–114. doi: 10.1111/j.1467-985X.2005.00389.x

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27,* 169–190. doi: 10.1017/s0140525x04000056

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36*, 329–347. doi:10.1017/S0140525X12001495

Plummer, M. (2008). Penalized loss functions for Bayesian model comparison. *Biostatistics, 9,* 523–539. doi: 10.1093/biostatistics/kxm049

Reitter, D., Keller, F., & Moore, J. D. (2011). A computational cognitive model of syntactic priming. *Cognitive Science, 35*, 587–637. doi:10.1111/j.1551-6709.2010.01165.x

Reitter, D., & Moore, J. D. (2014). Alignment and task success in spoken dialogue. *Journal of Memory and Language, 76*, 29–46. doi:10.1016/j.jml.2014.05.008

Schuurman, N. K., Grasman, R. P. P. P., & Hamaker, E. L. (2016). A comparison of inverse-Wishart prior specifications for covariance matrices in multilevel autoregressive models. *Multivariate Behavioral Research, 51,* 185–206. doi: 10.1080/00273171.2015.1065398

Skrondal, A. & Rabe-Hesketh, S. (2003). Multilevel logistic regression for polytomous data and rankings. *Psychometrika, 68,* 267–287. doi:10.1007/BF02294801

Song, H., & Zhang, Z. (2014). Analyzing multiple multivariate time series data using multilevel dynamic factor models. *Multivariate Behavioral Research, 49,* 67–77. doi:10.1080/00273171.2013.851018

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B, 64,* 583–616. doi:10.1111/1467-9868.00353

Spiegelhalter, D. J., Thomas, A., Best, N. G., & Lunn, D. (2003). WinBUGS user manual. Cambridge, UK: MRC Biostatistics Unit, Institute of Public Health. Retrieved from http://www.mrcbsu.cam.ac.uk/bugs/winbugs/manual14.pdf

Stave, M., Smolek, A., & Kapatsinski, V. (2013). *Inductive bias against stem changes as perseveration: Experimental evidence for an articulatory approach to output-output faithfulness.* In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 35, No. 35).

Thomas, K. & Clifford, S. (2017). Validity and mechanical turk: An assessment of exclusion methods and interactive experiments. *Computers in Human Behavior, 77,* 184–197. doi:10.1016/j.chb.2017.08.038.

Weiss, D. J., & Wark, J. (2009). Hysteresis effects in a motor task with cotton-top tamarins (Sanguinus oedipus). *Journal of Experimental Psychology: Animal Behavior Processes, 35,* 427–433. doi: 10.1037/a0013964

Appendix A: OpenBUGS Code for a Markov Mixed-Effect Multinomial Logistic Regression Model

The syntax file below is used to fit a Markov Mixed-Effect Multinomial Logistic Regression Model from Equations 4 and 5.

```
model{
#model specification
for (r in 1:R){
   y[r] ~ dcat(p[r,1:3])
}
for (r in 1:R){
   num[r,1] <- 1
for (k in 2:3) {
   num[r,k] <- exp(gamma1[k]*equals(dummy1[r],1)+theta0[person[r],k]*equals(dummy1[r],1)+beta0[item[r],k]*equals(dummy1[r],1)+
   gamma[k]*equals(dummy[r],1)+lambda2[k]*c1lag1[r]*equals(dummy[r],1)+lambda3[k]*c2lag1[r]*equals(dummy[r],1)+
   lambda2.R[person[r],k]*c1lag1[r]*equals(dummy[r],1)+lambda3.R[person[r],k]*c2lag1[r]*equals(dummy[r],1)+
   theta[person[r],k]*equals(dummy[r],1)+beta[item[r],k]*equals(dummy[r],1))
}}
for (r in 1:R){
   pdsum[r]<-sum(num[r,1:3])
for (k in 1:3) {
   p[r,k] <- num[r,k]/pdsum[r]
}}
#priors
for (k in 2:3) {
   gamma1[k] ~ dnorm(0,0.001)
   gamma[k] ~ dnorm(0,0.001)
   delta2[k] ~ dnorm(0,0.001)
   delta3[k] ~ dnorm(0,0.001)
}
#transformation
for (k in 2:3) {
   OR.lambda2[k] <- exp(lambda2[k])
   OR.lambda3[k] <- exp(lambda3[k])
}
#transformation
for (j in 1:J){
   lambda22[j] <- lambda2[2] + lambda2.R[j,2]
   lambda32[j] <- lambda2[3] + lambda2.R[j,3]
   lambda23[j] <- lambda3[2] + lambda3.R[j,2]
   lambda33[j] <- lambda3[3] + lambda3.R[j,3]
}
#transformation
for (j in 1:J){
   theta[j,2] <- theta1[j,1]
   theta[j,3] <- theta1[j,2]
   lambda2.R[j,2] <- theta1[j,3]
   lambda2.R[j,3] <- theta1[j,4]
   lambda3.R[j,2] <- theta1[j,5]
   lambda3.R[j,3] <- theta1[j,6]
}
#priors
for (i in 1:I){
   beta0[i,2:3]  ~ dmnorm(mu.be0[1:2], R.be0[1:2,1:2])
   beta[i,2:3]  ~ dmnorm(mu.be[1:2], R.be[1:2,1:2])
}
#priors
for (j in 1:J){
   theta0[j,2:3] ~ dmnorm(mu.th0[1:2], R.th0[1:2,1:2])
   theta1[j,1:6] ~ dmnorm(mu.th[1:6], R.th[1:6,1:6])
}
#constraints
mu.be0[1] <- 0;mu.be0[2] <- 0
```

```
mu.be[1] <- 0;mu.be[2] <- 0
mu.th0[1] <- 0;mu.th0[2] <- 0
mu.th[1] <- 0;mu.th[2] <- 0;mu.th[3] <- 0;mu.th[4] <- 0;mu.th[5] <- 0;mu.th[6] <- 0
#hyper-priors; transformation
R.be0[1:2,1:2] ~ dwish(Omega.be0[1:2,1:2], 3)
IR.be0[1:2,1:2] <- inverse(R.be0[1:2,1:2])

R.be[1:2,1:2] ~ dwish(Omega.be[1:2,1:2], 3)
IR.be[1:2,1:2] <- inverse(R.be[1:2,1:2])

R.th0[1:2,1:2] ~ dwish(Omega.th0[1:2,1:2], 3)
IR.th0[1:2,1:2] <- inverse(R.th0[1:2,1:2])

R.th[1:6,1:6] ~ dwish(Omega.th[1:6,1:6], 7)
IR.th[1:6,1:6] <- inverse(R.th[1:6,1:6])
}
```