Nested Subject-Verb Dependencies

Yair

12/18/2019

Contents

Preparing the data

Variables:

- nested: 1 if center embedding (objrel or objrel-nounpp), 0 mental embedding (SR or LR)
- long: 1 if LR (objrel-nounpp or mental embedding LR), 0 SR (objrel, mental embedding SR)
- violation_position: "inner" or "outer" - verb on which the violation occurred (mental embedding has only "inner" - verb in embedded clause)
- congruent_subjects: 1 if the two first subject nouns agree on number, else 0.
- number_v2: "singular" or "plural", based on the number of the second noun.
- congruent_attractor: 1 if the two last nouns agree on number, else 0.

For example, the last three variables defines: 'SSS' = (1, 'singular', 1), 'SSP' = (1, 'singular', 0), 'SPS' = (0, 'plural', 0), etc

Note that when long=0 (i.e., no attractor noun), then there are only four conditions (SS, SP, PS, PP), and therefore congruent_attractor='NA': 'SS'=(1, 'singular', 'NA'), etc.

```r
Data <- read.csv(file="../../Paradigm/Results/dataframe_results_all_trials.csv", header=TRUE, sep=",")
Data <- subset(Data, trial_type == "Violation") # Take only trials in which there was a violation
Data <- subset(Data, violation_position != "other") # Take only trials in which the violation was on th
Data <- subset(Data, valid_answer != "REJECTED") # remove from it rejected trials (in which subject did

# Remove the following columns (which will not be analyzed):
Data$violation_type <- NULL
Data$correct_wrong <- NULL
Data$block <- NULL
Data$trial_num <- NULL
Data$RT <- NULL
```

```
Data$slide_num_of_viol <- NULL

# Define 3 new binary columns that will classify all conditions ('SSS', 'SSP', 'SPS'...) based on wheth
Data$congruent_subjects <- ifelse(Data$condition == 'SSS' | Data$condition == 'SSP' | Data$condition ==
Data$number_v2 <- ifelse(Data$condition == 'SSS' | Data$condition == 'SSP' | Data$condition == 'PSS' | 
Data$congruent_attractor <- ifelse(Data$condition == 'SSS' | Data$condition == 'PSS' | Data$condition ==
Data$correct_wrong <- NULL
Data$condition <- NULL


# Define the main fixed variables (nested and long), based on the type of syntactic structure.
Data$nested <- ifelse(Data$sentence_type == "objrel" | Data$sentence_type == "objrel_nounpp", 1, 0)
Data$long <- ifelse(Data$sentence_type == "embedding_mental_LR" | Data$sentence_type == "objrel_nounpp"
Data$sentence_type <- NULL
Data$trial_type <- NULL
Data$valid_answer <- ifelse(Data$valid_answer == "CORRECT", 1, 0)
Data[] <- lapply(Data, function(x) if(is.factor(x)) factor(x) else x) # Remove all empty level in dataf
Data[] <- lapply(Data, factor) # change all variables to 'factor' type.
Data$valid_answer <- as.numeric(Data$valid_answer)-1

str(Data)
```

```
## 'data.frame':    7325 obs. of  8 variables:
##  $ subject           : Factor w/ 42 levels "21","22","23",..: 1 1 1 1 1 1 1 1 1 1 1 ...
##  $ violation_position : Factor w/ 2 levels "inner","outer": 1 1 1 1 1 1 1 1 1 1 1 ...
##  $ valid_answer      : num  1 1 1 1 0 1 1 1 1 0 ...
##  $ congruent_subjects : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 2 ...
##  $ number_v2         : Factor w/ 2 levels "plural","singular": 1 1 1 1 1 1 1 1 1 1 1 ...
##  $ congruent_attractor: Factor w/ 3 levels "0","1","NA": 2 2 2 2 2 1 1 1 1 1 ...
##  $ nested            : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 1 ...
##  $ long              : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 2 ...
```

```
summary(Data)
```

```
##      subject      violation_position  valid_answer    congruent_subjects
##  24     : 180    inner:4887          Min.   :0.0000   0:3657
##  28     : 180    outer:2438          1st Qu.:1.0000   1:3668
##  30     : 180                        Median :1.0000
##  34     : 180                        Mean   :0.8262
##  46     : 180                        3rd Qu.:1.0000
##  52     : 180                        Max.   :1.0000
##  (Other):6245
##     number_v2    congruent_attractor nested    long
##  plural  :4273   0 :2448             0:2442    0:2432
##  singular:3052   1 :2445             1:4883    1:4893
##                  NA:2432
##
##
##
##
```

Plot error rates

```
bysuj = with(Data, aggregate(valid_answer, list(subject=subject, nested=nested, long=long, violation_po
bysuj$error <- 1 - bysuj$x

par(mfrow=c(2,2))
for (n in 0:1)  {
  for (l in 0:1)  {
    struct = ifelse(n=="0", ifelse(l=="0", "Mental_SR", "Mental_LR"), ifelse(l=="0", "Objrel_SR", "Objr
    curr_bysuj = subset(Data, long == l & nested == n)
    curr_bysuj_mean = with(curr_bysuj, tapply(valid_answer, list(congruent_subjects=congruent_subjects,
    curr_bysuj_mean <- 1 - curr_bysuj_mean
    barplot(curr_bysuj_mean, col=c("red", "darkblue"), main = struct, ylim=c(0,0.6), names.arg = c("Inn
    legend("topright", c("Incongruent", "Congruent"), fill=c("red", "darkblue"))
  }
}
```
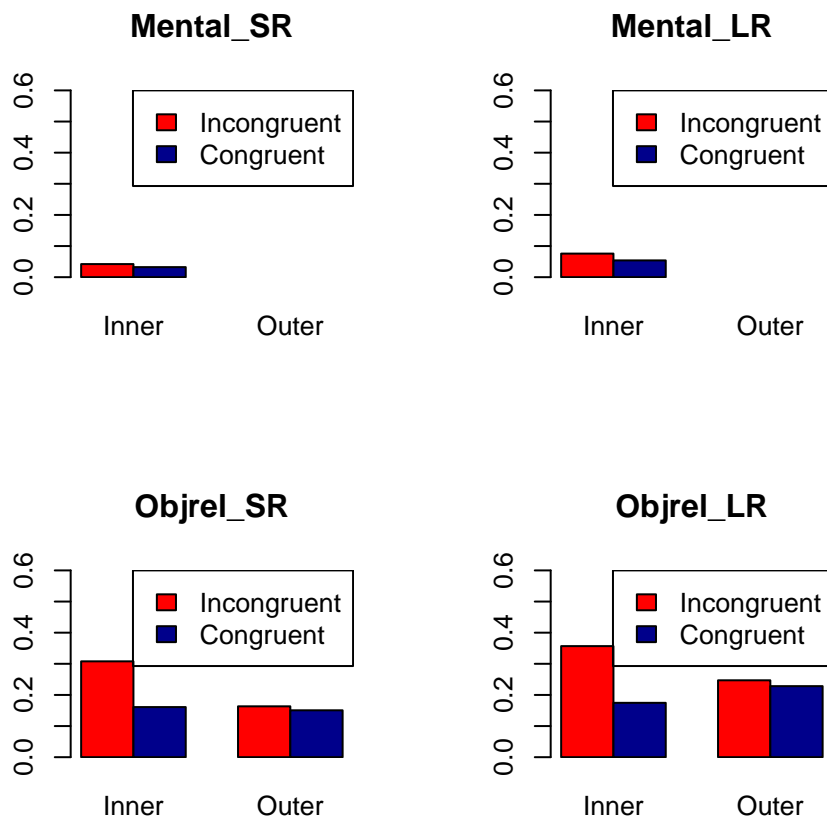

ANOVA: nested * long * subjects-congruency (inner verb only)

```
bysuj_inner <- subset(bysuj, violation_position=="inner")
with(bysuj_inner, interaction.plot(long, nested, error, ylim=c(0, 0.5)))
```
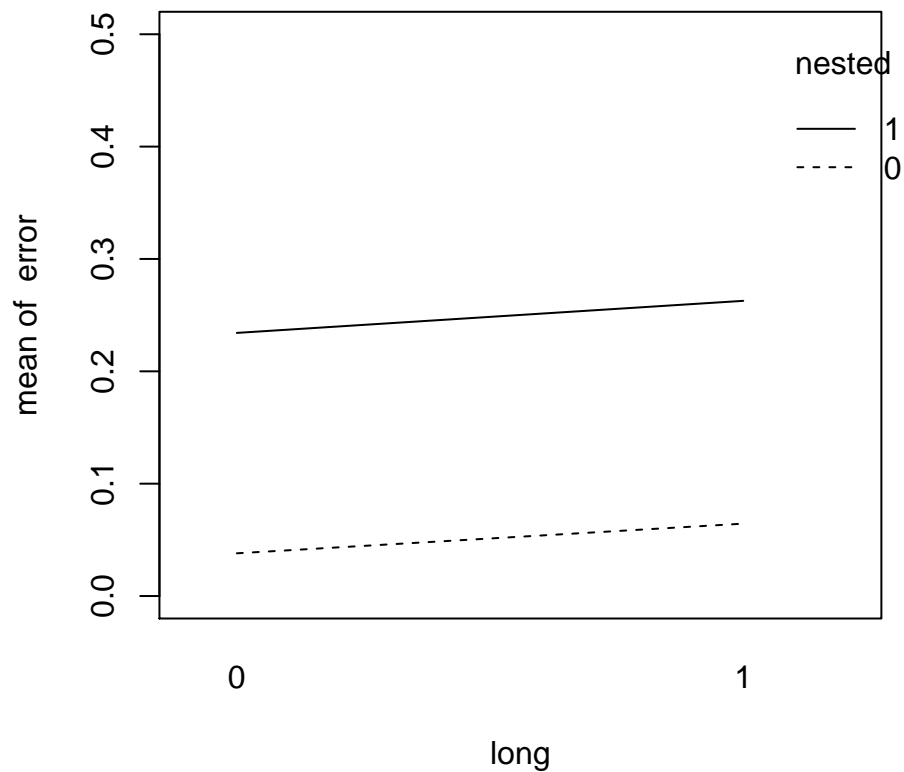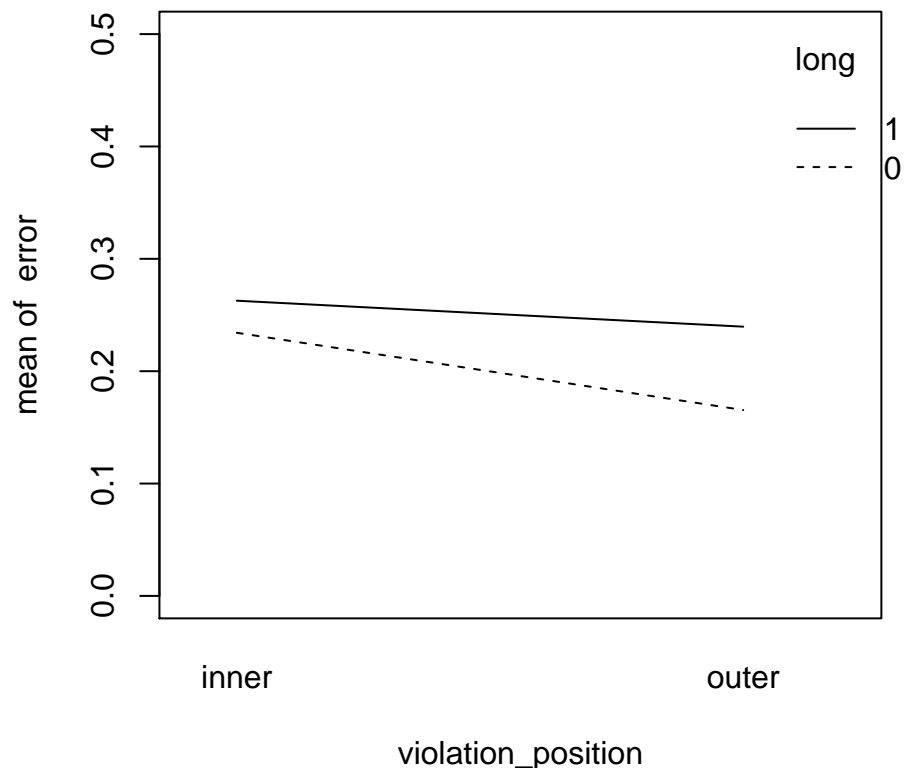
```
anov = aov(error ~ nested * long * congruent_subjects + Error(subject/(nested*long*congruent_subjects))
summary(anov)
```

```
##
## Error: subject
##            Df Sum Sq Mean Sq F value Pr(>F)
## Residuals 41  1.017  0.0248
##
## Error: subject:nested
##            Df Sum Sq Mean Sq F value Pr(>F)
## nested      1  3.268   3.268   233.8 <2e-16 ***
## Residuals 41  0.573   0.014
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Error: subject:long
##            Df Sum Sq Mean Sq F value Pr(>F)
## long        1 0.0633 0.06333   3.837  0.057 .
## Residuals 41 0.6768 0.01651
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Error: subject:congruent_subjects
##                    Df Sum Sq Mean Sq F value   Pr(>F)
## congruent_subjects  1 0.6638  0.6638   47.04 2.61e-08 ***
## Residuals          41 0.5787  0.0141
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Error: subject:nested:long
```

4

```
##              Df Sum Sq  Mean Sq F value Pr(>F)
## nested:long   1 0.0001 0.000094   0.008  0.929
## Residuals    41 0.4770 0.011634
##
## Error: subject:nested:congruent_subjects
##                             Df Sum Sq Mean Sq F value   Pr(>F)
## nested:congruent_subjects  1 0.4657  0.4657   41.93 9.12e-08 ***
## Residuals                 41 0.4554  0.0111
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Error: subject:long:congruent_subjects
##                             Df Sum Sq  Mean Sq F value Pr(>F)
## long:congruent_subjects    1 0.0156 0.015593   1.725  0.196
## Residuals                 41 0.3705 0.009037
##
## Error: subject:nested:long:congruent_subjects
##                                  Df  Sum Sq  Mean Sq F value Pr(>F)
## nested:long:congruent_subjects  1 0.00178 0.001782    0.25   0.62
## Residuals                      41 0.29236 0.007131
```

ANOVA: violation-position * long * subjects-congruency (objrel only)

```r
bysuj_nested <- subset(bysuj, nested=="1")
with(bysuj_nested, interaction.plot(violation_position, long, error, ylim=c(0, 0.5)))
```

```
anov = aov(error ~ long * congruent_subjects*violation_position + Error(subject/(long*congruent_subjects
summary(anov)
```

```
##
## Error: subject
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals 41  2.016 0.04917
##
## Error: subject:long
##           Df Sum Sq Mean Sq F value  Pr(>F)
## long       1 0.2215 0.22146   11.38 0.00163 **
## Residuals 41 0.7978 0.01946
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Error: subject:congruent_subjects
##                   Df Sum Sq Mean Sq F value   Pr(>F)
## congruent_subjects  1 0.6584  0.6584   43.47 6.21e-08 ***
## Residuals          41 0.6211  0.0151
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Error: subject:violation_position
##                   Df Sum Sq Mean Sq F value Pr(>F)
## violation_position  1 0.1772 0.17718   5.028 0.0304 *
## Residuals          41 1.4447 0.03524
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Error: subject:long:congruent_subjects
##                        Df Sum Sq Mean Sq F value Pr(>F)
## long:congruent_subjects  1 0.0113 0.01125   0.921  0.343
## Residuals               41 0.5008 0.01221
##
## Error: subject:long:violation_position
##                        Df Sum Sq Mean Sq F value Pr(>F)
## long:violation_position  1 0.0438 0.04379   2.619  0.113
## Residuals               41 0.6855 0.01672
##
## Error: subject:congruent_subjects:violation_position
##                                     Df Sum Sq Mean Sq F value   Pr(>F)
## congruent_subjects:violation_position  1 0.4702  0.4702   18.69 9.57e-05 ***
## Residuals                             41 1.0315  0.0252
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Error: subject:long:congruent_subjects:violation_position
##                                          Df Sum Sq  Mean Sq F value Pr(>F)
## long:congruent_subjects:violation_position  1 0.0037 0.003721   0.369  0.547
## Residuals                                  41 0.4133 0.010081
```

GLMM: nested * long * subjects-congruency (inner verb only)

```
glmm_with_random <- glmer(valid_answer ~ nested * long * congruent_subjects + (1 | subject), data=Data,
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.0110709 (tol = 0.001, component 1)
```

```
summary(glmm_with_random)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
## Family: binomial  ( logit )
## Formula: valid_answer ~ nested * long * congruent_subjects + (1 | subject)
##    Data: Data
##
##      AIC      BIC   logLik deviance df.resid
##   6157.5   6219.6  -3069.7   6139.5     7316
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -7.2925  0.1907  0.3592  0.5126  1.0844
##
## Random effects:
##  Groups  Name        Variance Std.Dev.
##  subject (Intercept) 0.1869   0.4324
## Number of obs: 7325, groups:  subject, 42
##
## Fixed effects:
##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                     3.21068    0.25821  12.434  < 2e-16 ***
## nested1                        -1.98837    0.26301  -7.560 4.03e-14 ***
## long1                          -0.62835    0.28246  -2.225   0.0261 *
## congruent_subjects1             0.27953    0.37766   0.740   0.4592
## nested1:long1                   0.27961    0.29992   0.932   0.3512
## nested1:congruent_subjects1     0.24878    0.39926   0.623   0.5332
## long1:congruent_subjects1       0.08825    0.43001   0.205   0.8374
## nested1:long1:congruent_subjects1 -0.05854  0.45686  -0.128   0.8980
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) nestd1 long1  cngr_1 nst1:1 ns1:_1 ln1:_1
## nested1    -0.915
## long1      -0.852  0.836
## cngrnt_sbj1 -0.639  0.628  0.586
## nestd1:lng1  0.802 -0.877 -0.942 -0.552
## nstd1:cng_1  0.606 -0.662 -0.554 -0.946  0.582
## lng1:cngr_1  0.563 -0.553 -0.661 -0.880  0.623  0.833
## nstd1:l1:_1 -0.530  0.580  0.622  0.829 -0.660 -0.875 -0.942
## convergence code: 0
## Model failed to converge with max|grad| = 0.0110709 (tol = 0.001, component 1)
```

```
glmm_wo_random <- glm(valid_answer ~ nested * long * congruent_subjects, data=Data, family="binomial")
summary(glmm_wo_random)
```

```
## 
## Call:
## glm(formula = valid_answer ~ nested * long * congruent_subjects,
##     family = "binomial", data = Data)
## 
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6207   0.2932   0.5821   0.6706   0.8475
## 
## Coefficients:
##                                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)                        3.12521    0.24776  12.614  < 2e-16 ***
## nested1                           -1.94938    0.26122  -7.463 8.48e-14 ***
## long1                             -0.62298    0.28077  -2.219   0.0265 *
## congruent_subjects1                0.27598    0.37513   0.736   0.4619
## nested1:long1                      0.28625    0.29767   0.962   0.3362
## nested1:congruent_subjects1        0.23779    0.39611   0.600   0.5483
## long1:congruent_subjects1          0.08658    0.42684   0.203   0.8393
## nested1:long1:congruent_subjects1 -0.06155    0.45288  -0.136   0.8919
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
##     Null deviance: 6766  on 7324  degrees of freedom
## Residual deviance: 6258  on 7317  degrees of freedom
## AIC: 6274
## 
## Number of Fisher Scoring iterations: 5
```

```
anova(glmm_with_random, glmm_wo_random)
```

```
## Data: Data
## Models:
## glmm_wo_random: valid_answer ~ nested * long * congruent_subjects
## glmm_with_random: valid_answer ~ nested * long * congruent_subjects + (1 | subject)
##                  Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## glmm_wo_random    8 6274.0 6329.1 -3129.0   6258.0
## glmm_with_random  9 6157.5 6219.6 -3069.8   6139.5 118.46      1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

GLMM: violation-position * long * subjects-congruency (objrel only)

```
glmm_with_random <- glmer(valid_answer ~ long * congruent_subjects*violation_position + (1 | subject),
summary(glmm_with_random)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: valid_answer ~ long * congruent_subjects * violation_position +
```

```
##     (1 | subject)
##    Data: Data
##
##      AIC     BIC   logLik deviance df.resid
##   6527.3  6589.4  -3254.7   6509.3     7316
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.9711  0.2936  0.3947  0.4902  0.9080
##
## Random effects:
##  Groups  Name        Variance Std.Dev.
##  subject (Intercept) 0.1678   0.4096
## Number of obs: 7325, groups:  subject, 42
##
## Fixed effects:
##                                                    Estimate Std. Error z value
## (Intercept)                                         1.60213    0.11314  14.161
## long1                                              -0.26381    0.11157  -2.364
## congruent_subjects1                                 0.69440    0.15134   4.588
## violation_positionouter                             0.08283    0.16491   0.502
## long1:congruent_subjects1                           0.07910    0.18096   0.437
## long1:violation_positionouter                      -0.26485    0.19422  -1.364
## congruent_subjects1:violation_positionouter        -0.59610    0.24707  -2.413
## long1:congruent_subjects1:violation_positionouter  -0.07252    0.29130  -0.249
##                                                    Pr(>|z|)
## (Intercept)                                        < 2e-16 ***
## long1                                               0.0181 *
## congruent_subjects1                                4.47e-06 ***
## violation_positionouter                             0.6155
## long1:congruent_subjects1                           0.6620
## long1:violation_positionouter                       0.1727
## congruent_subjects1:violation_positionouter         0.0158 *
## long1:congruent_subjects1:violation_positionouter   0.8034
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) long1  cngr_1 vltn_p ln1:_1 lng1:_ cn_1:_
## long1       -0.692
## cngrnt_sbj1 -0.509  0.517
## vltn_pstntr -0.468  0.474  0.350
## lng1:cngr_1  0.427 -0.617 -0.836 -0.293
## lng1:vltn_p  0.397 -0.575 -0.297 -0.849  0.355
## cngrnt_s1:_  0.312 -0.317 -0.613 -0.668  0.513  0.567
## lng1:cn_1:_ -0.265  0.383  0.520  0.566 -0.622 -0.667 -0.848
```

```r
glmm_wo_random <- glm(valid_answer ~ long * congruent_subjects*violation_position, data=Data, family="b
summary(glmm_wo_random)
```

```
##
## Call:
## glm(formula = valid_answer ~ long * congruent_subjects * violation_position,
##     family = "binomial", data = Data)
##
```

```
## Deviance Residuals:
##     Min      1Q  Median      3Q     Max
## -2.1593  0.4522  0.5714  0.6967  0.7531
##
## Coefficients:
##                                                    Estimate Std. Error z value
## (Intercept)                                         1.54846    0.09241  16.757
## long1                                              -0.25620    0.11032  -2.322
## congruent_subjects1                                 0.68060    0.15019   4.532
## violation_positionouter                             0.08493    0.16325   0.520
## long1:congruent_subjects1                           0.07396    0.17952   0.412
## long1:violation_positionouter                      -0.26214    0.19206  -1.365
## congruent_subjects1:violation_positionouter        -0.58422    0.24488  -2.386
## long1:congruent_subjects1:violation_positionouter  -0.06702    0.28846  -0.232
##                                                    Pr(>|z|)
## (Intercept)                                         < 2e-16 ***
## long1                                                0.0202 *
## congruent_subjects1                                5.85e-06 ***
## violation_positionouter                              0.6029
## long1:congruent_subjects1                            0.6803
## long1:violation_positionouter                        0.1723
## congruent_subjects1:violation_positionouter          0.0170 *
## long1:congruent_subjects1:violation_positionouter    0.8163
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 6766.0  on 7324  degrees of freedom
## Residual deviance: 6619.3  on 7317  degrees of freedom
## AIC: 6635.3
##
## Number of Fisher Scoring iterations: 4
```

```
anova(glmm_with_random, glmm_wo_random)
```

```
## Data: Data
## Models:
## glmm_wo_random: valid_answer ~ long * congruent_subjects * violation_position
## glmm_with_random: valid_answer ~ long * congruent_subjects * violation_position +
## glmm_with_random:     (1 | subject)
##                   Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## glmm_wo_random     8 6635.3 6690.5 -3309.7   6619.3
## glmm_with_random   9 6527.3 6589.4 -3254.7   6509.3 109.99      1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```