M21 LDT ERP N250

Joanna Morris

2025-06-22

Contents

Set parameters

Set chunk parameters

Load libraries

```
library(tidyverse)
library(ggeffects)
library(lme4)
library(afex)
library(gridExtra)
library(emmeans)
library(effectsize)
library(performance)
library(cowplot)  # for use with `plot_grid(x,x,ncol = x)` function
library(e1071) # for use with `skewness()` function
```

Set ggplot parameters

```
theme_set(theme_classic() +
            theme(legend.position = "bottom",
                  axis.text=element_text(size=10),
                  axis.title=element_text(size=9)))

# Define a custom color palette
my_palette <- c("#A6CEE3",  "#FB9A99")
my_palette_2 <- c( "#1F78B4","#E31A1C" )
my_palette_3 <- c("#A6CEE3","#1F78B4","#FB9A99","#E31A1C")


# Create a function to apply this palette
scale_color_custom <- function() {
  scale_color_manual(values = my_palette_2)
}

scale_fill_custom <- function() {
  scale_fill_manual(values = my_palette_2)
}
```

Define standard error of the mean function

```
sem <- function(x) sd(x)/sqrt(length(x))
```

1 N400

1.1 Load and format data files

Now we extract `SubjID` from the `ERPset` column

We then join the ERP data, and language into a single data frame

Divide into word, non-word and difference wave dataframes

```
n400_1_words <- n400_1 |> filter(bini %in% c(1:2))
n400_1_words_b <- n400_1 |> filter(bini %in% c(9:12))
n400_1_nonwords <- n400_1 |> filter(bini %in% c(3:6))
```

Then we do some more formatting and cleanup of the dataframes.We create separate columns, one for each independent variable (anteriority, laterality, morphological family size). To do this we have to use `seperate` function from the `stringr` package. Run `vignette("programming", package = "dplyr")` to see more about `tidy-selection` and `tidy-evaluation`.

Now we need to extract just the bins and channels that we intend to analyse. For this analysis we will use 9 channels: F3, Fz, F4, C3, Cz, C4, P3, Pz, P4 . We will use the`mutate` function from the `dplyr` package along with the `case_when` function. The `case_when` function is a sequence of two-sided formulas. The left hand side determines which values match this case. The right hand side provides the replacement value.

1.2 Real Word Data

1.2.1 Compute the ANOVA

```
anova_model_1a <- mixed(
  value ~ orthographic_sensitivity * family_size * base_freq  +
    laterality * anteriority  +  # Nuisance variables
    (1 | SubjID),
  data = n400_1_words_b,
  method = "KR")  # Kenward-Roger approximation for accurate F-tests
# Print ANOVA results
anova_model_1a
```

```
|| Mixed Model Anova Table (Type 3 tests, KR-method)
||
|| Model: value ~ orthographic_sensitivity * family_size * base_freq +
|| Model:     laterality * anteriority + (1 | SubjID)
|| Data: n400_1_words_b
||                                          Effect      df          F p.value
|| 1                          orthographic_sensitivity   1, 59     4.83 *    .032
|| 2                                      family_size 1, 2121  11.20 ***   <.001
|| 3                                        base_freq 1, 2121     3.09 +    .079
|| 4                                       laterality 2, 2121    4.75 **    .009
|| 5                                      anteriority 2, 2121 104.31 ***   <.001
|| 6            orthographic_sensitivity:family_size 1, 2121       0.01    .932
|| 7              orthographic_sensitivity:base_freq 1, 2121       0.00    .973
|| 8                          family_size:base_freq 1, 2121  23.41 ***   <.001
|| 9                        laterality:anteriority 4, 2121       0.77    .542
|| 10 orthographic_sensitivity:family_size:base_freq 1, 2121       0.16    .692
|| ---
|| Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '+' 0.1 ' ' 1
```

```
# Partial Eta Squared
# Extract effect sizes from your ANOVA model
eta_squared(anova_model_1a , partial = TRUE)
```

```
|| # Effect Size for ANOVA (Type III)
||
|| Parameter                                   | Eta2 (partial) |       95% CI
|| ------------------------------------------------------------------------------
|| orthographic_sensitivity                    |           0.08 | [0.00, 1.00]
|| family_size                                 |       5.25e-03 | [0.00, 1.00]
|| base_freq                                   |       1.46e-03 | [0.00, 1.00]
|| laterality                                  |       4.46e-03 | [0.00, 1.00]
|| anteriority                                 |           0.09 | [0.07, 1.00]
|| orthographic_sensitivity:family_size        |       3.42e-06 | [0.00, 1.00]
|| orthographic_sensitivity:base_freq          |       5.56e-07 | [0.00, 1.00]
|| family_size:base_freq                       |           0.01 | [0.00, 1.00]
|| laterality:anteriority                      |       1.46e-03 | [0.00, 1.00]
|| orthographic_sensitivity:family_size:base_freq |       7.38e-05 | [0.00, 1.00]
||
|| - One-sided CIs: upper bound fixed at [1.00].
```

```
# Compute Marginal (fixed effects) and Conditional (fixed + random effects) R²
r2(anova_model_1a)
```

```
|| # R2 for Mixed Models
||
||   Conditional R2: 0.567
```

```
||      Marginal R2: 0.090
```

1.2.2 Significant Effects

Effect	df	F	p.value	
orthographic_sensitivity	1, 59	4.83 *	.032	0.08
family_size	1, 2121	11.20 ***	<.001	5.25e-03
base_freq	1, 2121	3.09 +	.079	1.46e-03
family_size:base_freq	1, 2121	4.75 **	.009	0.01

```
## `orthographic_sensitivity` main effect
# Get estimated marginal means for each level of complexity
pairs <- emmeans(anova_model_1a, pairwise ~ orthographic_sensitivity, adjust = "bonferroni", pbkrtest.limit = 6480)
pairs_df <- as.data.frame(pairs$contrasts)
cohensd <- as.data.frame(cohens_d(value ~ orthographic_sensitivity, data = n400_1_words_b))
(orthographic_sensitivity_contrasts_df <- bind_cols(pairs_df,cohensd))
```

```
||  contrast                          estimate       SE df t.ratio p.value
||  High Sensitivity - Low Sensitivity 1.496905 0.681261 59   2.197  0.0319
||   Cohens_d   CI    CI_low   CI_high
||  0.4066995 0.95 0.3218643 0.4914439
||
|| Results are averaged over the levels of: family_size, base_freq, laterality, anteriority
|| Degrees-of-freedom method: kenward-roger
```

```
(orthographic_sensitivity_means <- as.data.frame(pairs$emmeans))
```

```
||  orthographic_sensitivity     emmean        SE df   lower.CL  upper.CL
||  High Sensitivity          1.4730030 0.5010783 59  0.4703477 2.4756583
||  Low Sensitivity          -0.0239015 0.4615594 59 -0.9474798 0.8996767
||
|| Results are averaged over the levels of: family_size, base_freq, laterality, anteriority
|| Degrees-of-freedom method: kenward-roger
|| Confidence level used: 0.95
```

```
## `family_size` main effect
# Get estimated marginal means for each level of family size
pairs <- emmeans(anova_model_1a, pairwise ~ family_size, adjust = "bonferroni", pbkrtest.limit = 6480)
pairs_df <- as.data.frame(pairs$contrasts)
cohensd <- as.data.frame(cohens_d(value ~ family_size, data = n400_1_words_b))
(family_size_contrasts_df <- bind_cols(pairs_df,cohensd))
```

```
||  contrast                    estimate        SE   df t.ratio p.value
||  Large Family - Small Family 0.3572769 0.1067486 2121   3.347  0.0008
||    Cohens_d   CI     CI_low  CI_high
||  0.09544102 0.95 0.01173328 0.179127
||
|| Results are averaged over the levels of: orthographic_sensitivity, base_freq, laterality, anteriority
|| Degrees-of-freedom method: kenward-roger
```

```
(family_size_means <- as.data.frame(pairs$emmeans))
```

```
||  family_size     emmean        SE    df   lower.CL upper.CL
||  Large Family 0.9031892 0.3447868 61.93  0.2139551 1.592423
||  Small Family 0.5459123 0.3447868 61.93 -0.1433218 1.235146
||
|| Results are averaged over the levels of: orthographic_sensitivity, base_freq, laterality, anteriority
|| Degrees-of-freedom method: kenward-roger
|| Confidence level used: 0.95
```

```
## `base_freq` main effect
# Get estimated marginal means for each level of base frequency
pairs <- emmeans(anova_model_1a, pairwise ~ base_freq, adjust = "bonferroni", pbkrtest.limit = 6480)
pairs_df <- as.data.frame(pairs$contrasts)
cohensd <- as.data.frame(cohens_d(value ~ base_freq, data = n400_1_words_b))
(base_frequency_contrasts_df <- bind_cols(pairs_df,cohensd))
```

```
||  contrast                                estimate        SE   df t.ratio
||  High Base Frequency - Low Base Frequency 0.1877249 0.1067486 2121   1.759
||  p.value   Cohens_d   CI      CI_low   CI_high
||   0.0788 0.05008228 0.95 -0.03358577 0.1337389
||
|| Results are averaged over the levels of: orthographic_sensitivity, family_size, laterality, anteriority
|| Degrees-of-freedom method: kenward-roger
```

```
(base_frequency_means <- as.data.frame(pairs$emmeans))
```

```
||  base_freq              emmean        SE    df    lower.CL upper.CL
||  High Base Frequency 0.8184132 0.3447868 61.93  0.12917906 1.507647
||  Low Base Frequency  0.6306883 0.3447868 61.93 -0.05854584 1.319922
```

```
||
|| Results are averaged over the levels of: orthographic_sensitivity, family_size, laterality, anteriority
|| Degrees-of-freedom method: kenward-roger
|| Confidence level used: 0.95
```

```
# Test whether the interaction between family_size and base_freq improves model fit
reduced_model_int <- update(anova_model_1a,
  . ~ . - family_size:base_freq - orthographic_sensitivity:family_size:base_freq)

anova(anova_model_1a, reduced_model_int)
```

```
|| Data: data
|| Models:
|| reduced_model_int: value ~ orthographic_sensitivity + family_size + base_freq + laterality + anteriority + orthographic_sensitivity:family_size
|| anova_model_1a: value ~ orthographic_sensitivity * family_size * base_freq + laterality * anteriority + (1 | SubjID)
||                   npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
|| reduced_model_int   16 10551 10643 -5259.7    10519
|| anova_model_1a      18 10540 10643 -5252.1    10504 15.252  2  0.0004876 ***
|| ---
|| Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Custom contrasts for family_size × base_freq Interaction
pairs <- emmeans(anova_model_1a, pairwise ~ family_size * base_freq, adjust = "bonferroni", pbkrtest.limit = 6480)
(pairs_df <- as.data.frame(pairs$contrasts))
```

```
||  contrast                                                    estimate
||  Large Family High Base Frequency - Small Family High Base Frequency -0.1592380
||  Large Family High Base Frequency - Large Family Low Base Frequency  -0.3287900
||  Large Family High Base Frequency - Small Family Low Base Frequency   0.5450018
||  Small Family High Base Frequency - Large Family Low Base Frequency  -0.1695520
||  Small Family High Base Frequency - Small Family Low Base Frequency   0.7042398
||  Large Family Low Base Frequency - Small Family Low Base Frequency    0.8737918
||         SE   df t.ratio p.value
||  0.1509653 2121  -1.055  1.0000
||  0.1509653 2121  -2.178  0.1771
||  0.1509653 2121   3.610  0.0019
||  0.1509653 2121  -1.123  1.0000
||  0.1509653 2121   4.665  <.0001
||  0.1509653 2121   5.788  <.0001
||
|| Results are averaged over the levels of: orthographic_sensitivity, laterality, anteriority
|| Degrees-of-freedom method: kenward-roger
|| P value adjustment: bonferroni method for 6 tests
```

```
selected_contrasts_basefrq <- pairs$contrasts[pairs_df$contrast %in% c("Large Family High Base Frequency - Large Family Low Base Frequency",
                                                                       "Small Family High Base Frequency - Small Family Low Base Frequency"),]
selected_contrasts_famsize <- pairs$contrasts[pairs_df$contrast %in% c("Large Family High Base Frequency - Small Family High Base Frequency",
                                                                       "Large Family Low Base Frequency - Small Family Low Base Frequency"), ]

selected_contrasts_basefrq_df <- as.data.frame(selected_contrasts_basefrq)  # Convert the emmGrid object to a dataframe
selected_contrasts_famsize_df <- as.data.frame(selected_contrasts_famsize)

cohensd_hi_basefrq <- as.data.frame(cohens_d(value ~ family_size,
                                     data = subset(n400_1_words_b, base_freq == "High Base Frequency")))
cohensd_lo_basefrq <- as.data.frame(cohens_d(value ~ family_size,
                                      data = subset(n400_1_words_b, base_freq == "Low Base Frequency")))
cohensd_lrg_famsize <- as.data.frame(cohens_d(value ~ base_freq,
                                       data = subset(n400_1_words_b, family_size == "Large Family")))
cohensd_sml_famsize <- as.data.frame(cohens_d(value ~ base_freq,
                                      data = subset(n400_1_words_b, family_size == "Small Family")))

cohensd_basefrq <- bind_rows(hi_basefrq = cohensd_hi_basefrq,
                             lo_basefrq = cohensd_lo_basefrq,
                             .id = "base_freq")

cohensd_famsize <- bind_rows(hifamsize = cohensd_lrg_famsize,
                             lo_sensi = cohensd_sml_famsize,
                             .id = "family_size")

(basefreq_contrasts_df <- bind_cols(selected_contrasts_basefrq_df,cohensd_basefrq))
```

```
||  contrast                                                    estimate
||  Large Family High Base Frequency - Large Family Low Base Frequency -0.3287900
||  Small Family High Base Frequency - Small Family Low Base Frequency  0.7042398
||         SE   df t.ratio p.value base_freq    Cohens_d   CI     CI_low   CI_high
||  0.1509653 2121  -2.178  0.0590 hi_basefrq -0.0407526 0.95 -0.1590535 0.0775669
||  0.1509653 2121   4.665  <.0001 lo_basefrq  0.2366583 0.95  0.1178928 0.3553166
||
|| Results are averaged over the levels of: orthographic_sensitivity, laterality, anteriority
|| Degrees-of-freedom method: kenward-roger
|| P value adjustment: bonferroni method for 2 tests
```

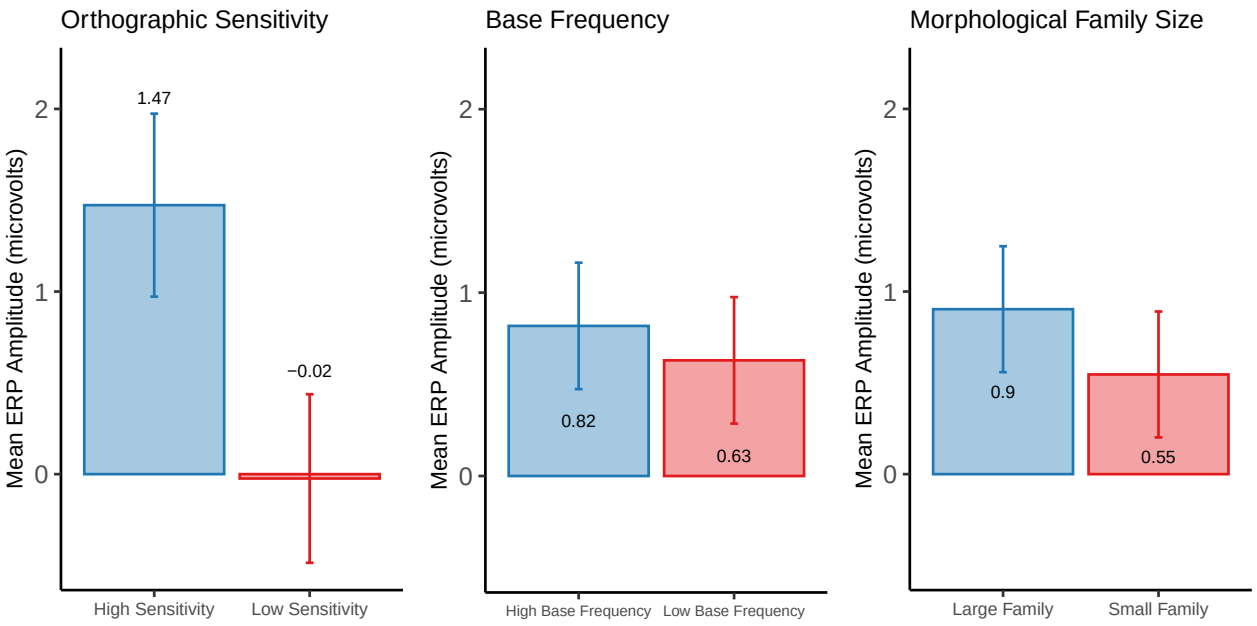
```
(family_size_contrasts_df <- bind_cols(selected_contrasts_famsize_df,cohensd_famsize))
```

```
||  contrast                                                          estimate
||  Large Family High Base Frequency - Small Family High Base Frequency -0.1592380
||  Large Family Low Base Frequency - Small Family Low Base Frequency    0.8737918
||         SE   df t.ratio p.value family_size   Cohens_d   CI      CI_low
||  0.1509653 2121  -1.055  0.5833 hifamsize   -0.0915178 0.95 -0.20985682
||  0.1509653 2121   5.788  <.0001 lo_sensi     0.1786143 0.95  0.06003985
||     CI_high
||  0.02686293
||  0.29710762
||
|| Results are averaged over the levels of: orthographic_sensitivity, laterality, anteriority
|| Degrees-of-freedom method: kenward-roger
|| P value adjustment: bonferroni method for 2 tests
```

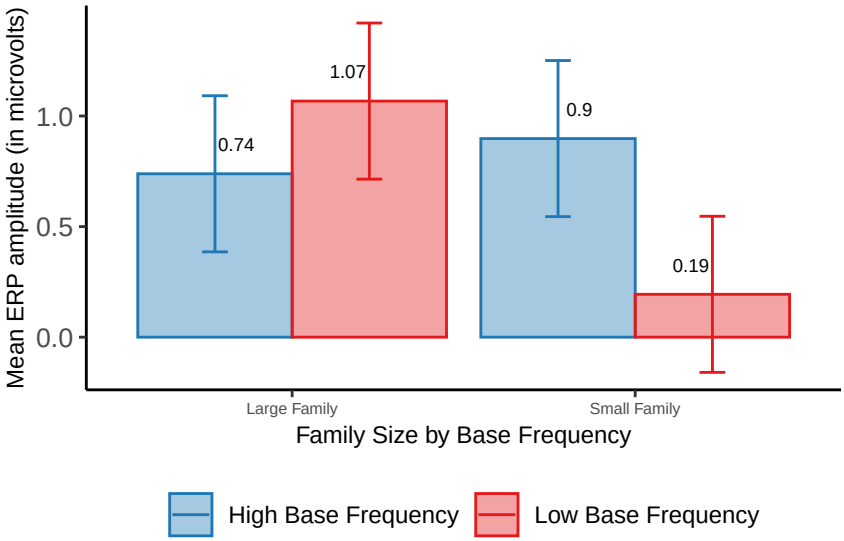
```
(family_size.basefreq_means <- as.data.frame(pairs$emmeans))
```

```
||  family_size  base_freq             emmean        SE df   lower.CL  upper.CL
||  Large Family High Base Frequency 0.7387942 0.3529527 68  0.0344884 1.4431000
||  Small Family High Base Frequency 0.8980322 0.3529527 68  0.1937264 1.6023380
||  Large Family Low Base Frequency  1.0675842 0.3529527 68  0.3632784 1.7718900
||  Small Family Low Base Frequency  0.1937924 0.3529527 68 -0.5105134 0.8980982
||
|| Results are averaged over the levels of: orthographic_sensitivity, laterality, anteriority
|| Degrees-of-freedom method: kenward-roger
|| Confidence level used: 0.95
```

```
p1 <- ggplot(orthographic_sensitivity_means, aes(x = orthographic_sensitivity, y = emmean,
                                                 fill = orthographic_sensitivity, colour = orthographic_sensitivity)) +
  geom_bar(stat = "identity", position = position_dodge(), alpha = .4) +
  geom_errorbar(aes(ymin = emmean - SE, ymax = emmean + SE),
                width = 0.05, position = position_dodge(0.9)) +
  ylab("Mean ERP Amplitude (microvolts)") +
  geom_text(aes(label = round(emmean, digits = 2)), colour = "black", size = 2.5, vjust = -8,
            position = position_dodge(.9)) +
  coord_cartesian(ylim = c(-.5, 2.2)) +
  scale_color_custom() +
  scale_fill_custom() +
  labs(title = "Orthographic Sensitivity") +
  theme( plot.title = element_text(size = 10),
         legend.position = "none",
         axis.title.x = element_blank(),
         axis.text.x = element_text(size = 7))


p2 <- ggplot(base_frequency_means, aes(x = base_freq, y = emmean, fill = base_freq, colour = base_freq)) +
  geom_bar(stat = "identity", position = position_dodge(), alpha = .4) +
  geom_errorbar(aes(ymin = emmean - SE, ymax = emmean + SE),
                width = 0.05, position = position_dodge(0.9)) +
  ylab("Mean ERP Amplitude (microvolts)") +
  geom_text(aes(label = round(emmean, digits = 2)), colour = "black", size = 2.5, vjust = 8,
            position = position_dodge(.9)) +
  coord_cartesian(ylim = c(-.5, 2.2)) +
  scale_color_custom() +
  scale_fill_custom() +
  labs(title = "Base Frequency") +
  theme( plot.title = element_text(size = 10),
         legend.position = "none",
         axis.title.x = element_blank(),
         axis.text.x = element_text(size = 6))


p3 <- ggplot(family_size_means, aes(x = family_size, y = emmean, fill = family_size, colour = family_size)) +
  geom_bar(stat = "identity", position = position_dodge(), alpha = .4) +
  geom_errorbar(aes(ymin = emmean - SE, ymax = emmean + SE),
                width = 0.05, position = position_dodge(0.9)) +
  ylab("Mean ERP Amplitude (microvolts)") +
  geom_text(aes(label = round(emmean, digits = 2)), colour = "black", size = 2.5, vjust = 7,
            position = position_dodge(.9)) +
  coord_cartesian(ylim = c(-.5, 2.2)) +
  scale_color_custom() +
  scale_fill_custom() +
  labs(title = "Morphological Family Size" ) +
  theme( plot.title = element_text(size = 10),
         legend.position = "none",
         axis.title.x = element_blank(),
         axis.text.x = element_text(size = 7))

plot_grid(p1, p2, p3, ncol = 3)
```



Plot for `family_size` × `base_freq` Interaction



1.3 Nonword Data

1.3.1 Compute the ANOVA

```
# Fit the ANOVA/mixed model
anova_model_1b <- mixed(
  value ~ orthographic_sensitivity * family_size * complexity +
    laterality * anteriority  +  # Nuisance variables
    (1 | SubjID),
  data = n400_1_nonwords,
  method = "KR"  # Kenward-Roger approximation for accurate F-tests
)

# Print ANOVA results
anova_model_1b
```

```
|| Mixed Model Anova Table (Type 3 tests, KR-method)
||
|| Model: value ~ orthographic_sensitivity * family_size * complexity +
|| Model:     laterality * anteriority + (1 | SubjID)
|| Data: n400_1_nonwords
||                                         Effect      df          F p.value
```

```
|| 1                         orthographic_sensitivity   1, 59       1.83    .181
|| 2                                   family_size 1, 2121       0.04    .841
|| 3                                    complexity 1, 2121       0.00    .953
|| 4                                    laterality 2, 2121     4.22 *    .015
|| 5                                   anteriority 2, 2121 140.82 ***   <.001
|| 6            orthographic_sensitivity:family_size 1, 2121       0.70    .402
|| 7             orthographic_sensitivity:complexity 1, 2121       2.31    .129
|| 8                       family_size:complexity 1, 2121       1.35    .246
|| 9                       laterality:anteriority 4, 2121       0.40    .812
|| 10 orthographic_sensitivity:family_size:complexity 1, 2121       2.40    .121
|| ---
|| Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '+' 0.1 ' ' 1
```

```r
# Partial Eta Squared
# Extract effect sizes from your ANOVA model
eta_squared(anova_model_1b , partial = TRUE)
```

```
|| # Effect Size for ANOVA (Type III)
||
|| Parameter                                        | Eta2 (partial) |       95% CI
|| -------------------------------------------------------------------------------------
|| orthographic_sensitivity                         |           0.03 | [0.00, 1.00]
|| family_size                                      |       1.89e-05 | [0.00, 1.00]
|| complexity                                       |       1.64e-06 | [0.00, 1.00]
|| laterality                                       |       3.96e-03 | [0.00, 1.00]
|| anteriority                                      |           0.12 | [0.10, 1.00]
|| orthographic_sensitivity:family_size             |       3.31e-04 | [0.00, 1.00]
|| orthographic_sensitivity:complexity              |       1.09e-03 | [0.00, 1.00]
|| family_size:complexity                           |       6.36e-04 | [0.00, 1.00]
|| laterality:anteriority                           |       7.46e-04 | [0.00, 1.00]
|| orthographic_sensitivity:family_size:complexity |       1.13e-03 | [0.00, 1.00]
||
|| - One-sided CIs: upper bound fixed at [1.00].
```

```r
# Compute Marginal (fixed effects) and Conditional (fixed + random effects) R²
r2(anova_model_1b)
```

```
|| # R2 for Mixed Models
||
||   Conditional R2: 0.522
||      Marginal R2: 0.079
```

1.3.2 Significant Effects

No significant effects of any of the experimental variables were found.