M21 LDT ERP HC ORTHOGRAPIC SENSITIVITY

Joanna Morris

2025-10-04

Contents

Set parameters

Set chunk parameters

Load libraries

Set ggplot parameters

Define standard error of the mean function

1 Load and format data files

```
dir_path <- "CSV files"

erp_2 <- read_csv(file.path(dir_path, "m21_ldt_mea_200300_050050_1.csv"))
erp_4 <- read_csv(file.path(dir_path, "m21_ldt_mea_300500_050050_1.csv"))
dmg_lng_vsl <- read_csv(file.path(dir_path, "demo_lang_vsl_pca_hc.csv"))
```

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

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.

2 N250 Word Data

2.1 Nested ANOVA Model

```
#Fit ANOVA model
anova_model_n250_words_b <- mixed(
    value ~ Orthographic_Sensitivity * family_size * base_freq +
    (1 + family_size + base_freq | SubjID) +     # by-subject intercept + slopes
    (1 | SubjID:chlabel),                        # electrode nested within subject
  data   = n250_words_b,
  method = "KR"
)
anova_model_n250_words_b
```

```
|| Mixed Model Anova Table (Type 3 tests, KR-method)
||
|| Model: value ~ Orthographic_Sensitivity * family_size * base_freq +
|| Model:     (1 + family_size + base_freq | SubjID) + (1 | SubjID:chlabel)
|| Data: n250_words_b
||                                         Effect      df         F p.value
|| 1                       Orthographic_Sensitivity   1, 59      0.03    .854
|| 2                                    family_size   1, 59      1.07    .306
|| 3                                      base_freq   1, 59      1.12    .294
|| 4           Orthographic_Sensitivity:family_size   1, 59      0.09    .762
|| 5             Orthographic_Sensitivity:base_freq   1, 59      0.12    .734
|| 6                        family_size:base_freq 1, 1523 35.14 ***   <.001
|| 7 Orthographic_Sensitivity:family_size:base_freq 1, 1523      0.02    .884
|| ---
|| Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '+' 0.1 ' ' 1
```

```
m1 <- anova_model_n250_words_b$full_model    # Extract the lmer model
ranova(m1) # Run random effects comparison
```

```
|| ANOVA-like table for random-effects: Single term deletions
||
|| Model:
|| value ~ Orthographic_Sensitivity + family_size + base_freq + (1 + family_size + base_freq | SubjID) + (1 | SubjID:chlabel) + Orthographic_Sensi
||                                                    npar  logLik    AIC    LRT Df Pr(>Chisq)
|| <none>                                               16 -4489.4 9010.8
|| family_size in (1 + family_size + base_freq | SubjID)   13 -4803.0 9631.9 627.07  3  < 2.2e-16 ***
|| base_freq in (1 + family_size + base_freq | SubjID)     13 -4716.5 9459.0 454.13  3  < 2.2e-16 ***
|| (1 | SubjID:chlabel)                                    15 -4684.5 9399.0 390.18  1  < 2.2e-16 ***
|| ---
|| Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Extract effect sizes from your ANOVA model
eta_squared(anova_model_n250_words_b, partial = TRUE)
```

```
|| # Effect Size for ANOVA (Type III)
||
|| Parameter                                 | Eta2 (partial) |       95% CI
|| ------------------------------------------------------------------------------
|| Orthographic_Sensitivity                  |       5.82e-04 | [0.00, 1.00]
|| family_size                               |           0.02 | [0.00, 1.00]
|| base_freq                                 |           0.02 | [0.00, 1.00]
|| Orthographic_Sensitivity:family_size      |       1.56e-03 | [0.00, 1.00]
|| Orthographic_Sensitivity:base_freq        |       1.97e-03 | [0.00, 1.00]
|| family_size:base_freq                     |           0.02 | [0.01, 1.00]
|| Orthographic_Sensitivity:family_size:base_freq |       1.40e-05 | [0.00, 1.00]
||
|| - One-sided CIs: upper bound fixed at [1.00].
```

```
# Compute Marginal(fixed effects only) and Conditional(fixed + random effects) R²
r2(anova_model_n250_words_b)
```

```
|| # R2 for Mixed Models
||
||   Conditional R2: 0.786
||      Marginal R2: 0.008
```

2.2 Significant Effects

Effect	df	F	p.value	
family_size:base_freq	1, 1523	35.14 ***	<.001	6.76e-03

2.2.1 Main Effects

No significant main effects

2.2.2 Interactions

```
# `base_freq` x `family_size` interaction

# Estimated marginal means for the family_size × base_freq interaction
emm <- emmeans(anova_model_n250_words_b, ~ family_size * base_freq)

# Look at the table of estimated means
emm
```

```
||  family_size  base_freq           emmean    SE   df lower.CL upper.CL
||  Large Family High Base Frequency -0.919 0.284 60.4    -1.49   -0.351
||  Small Family High Base Frequency -0.829 0.352 59.9    -1.53   -0.125
||  Large Family Low Base Frequency  -0.327 0.292 60.3    -0.91    0.256
||  Small Family Low Base Frequency  -0.952 0.344 59.9    -1.64   -0.264
||
|| Results are averaged over the levels of: Orthographic_Sensitivity
|| Degrees-of-freedom method: kenward-roger
|| Confidence level used: 0.95
```

```
# Simple effects of family_size at each level of base_freq
contrast(emm, method = "pairwise", by = "base_freq", adjust = "holm")
```

```
|| base_freq = High Base Frequency:
||  contrast                   estimate    SE   df t.ratio p.value
||  Large Family - Small Family  -0.0895 0.266 65.5  -0.337  0.7375
||
|| base_freq = Low Base Frequency:
||  contrast                   estimate    SE   df t.ratio p.value
||  Large Family - Small Family   0.6246 0.266 65.5   2.350  0.0218
||
|| Results are averaged over the levels of: Orthographic_Sensitivity
|| Degrees-of-freedom method: kenward-roger
```

```
# Simple effects of base_freq at each level of family_size
contrast(emm, method = "pairwise", by = "family_size", adjust = "holm")
```

```
|| family_size = Large Family:
||  contrast                                estimate   SE df t.ratio p.value
||  High Base Frequency - Low Base Frequency   -0.592 0.23 68  -2.576  0.0122
||
|| family_size = Small Family:
||  contrast                                estimate   SE df t.ratio p.value
||  High Base Frequency - Low Base Frequency    0.122 0.23 68   0.532  0.5967
||
|| Results are averaged over the levels of: Orthographic_Sensitivity
|| Degrees-of-freedom method: kenward-roger
```

```
# Interaction contrasts (e.g., difference of differences)
contrast(emm, interaction = "pairwise", adjust = "holm")
```

```
||  family_size_pairwise        base_freq_pairwise                     estimate   SE   df t.ratio p.value
||  Large Family - Small Family High Base Frequency - Low Base Frequency   -0.714 0.12 1523  -5.928  <.0001
||
|| Results are averaged over the levels of: Orthographic_Sensitivity
|| Degrees-of-freedom method: kenward-roger
```

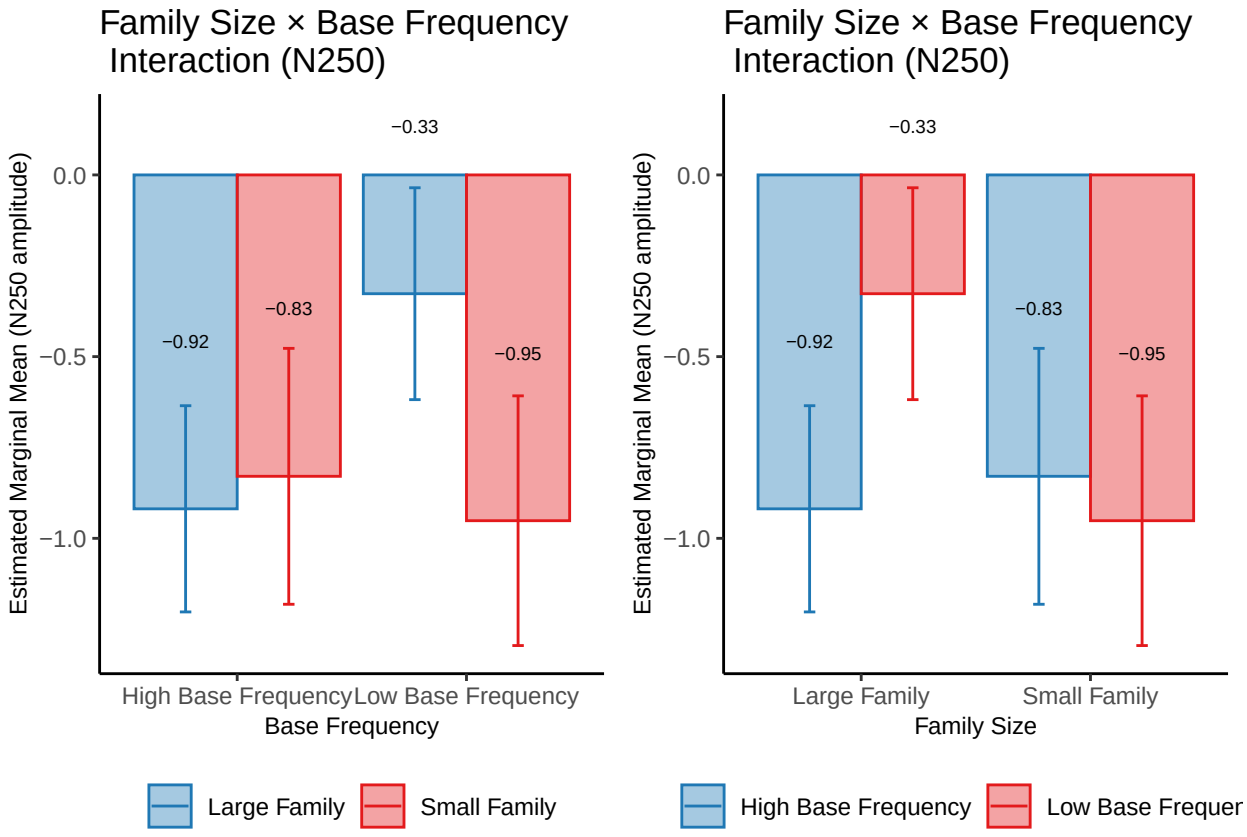
For large-family words, N250 amplitude is more negative when base frequency is high than when it is low. For small-family words, base frequency has little effect. For low-frequency bases, small-family words elicit more negative amplitudes than large-family words.

- At **High Base Frequency**: `Large` vs. `Small` family → no difference ($p = .74$). Family size doesn't matter when base frequency is high.
- Within **Small Family**: `High` vs. `Low` base frequency → not significant ($p = .60$). Small-family words are unaffected by base frequency.
- At **Low Base Frequency**: Large vs. Small family → significant difference ($p = .022$). Small-family words yield more negative amplitudes than large-family words, but only when base frequency is low.
- Within **Large Family**: High vs. Low base frequency → significant ($p = .012$). Large-family words show more negative amplitudes when their base frequency is high.

2.3 Plots

```
# Get estimated marginal means for the interaction
emm <- emmeans(anova_model_n250_words_b, ~ family_size * base_freq)

# Convert to data frame for plotting
emm_df <- as.data.frame(emm)

# Plot barchart
p1 <- ggplot(emm_df, aes(x = base_freq, y = emmean,
```

```
                  color = family_size, fill = family_size)) +
  geom_col(alpha = .4, position = position_dodge(0.9)) +
  geom_errorbar(aes(ymin = emmean - SE, ymax = emmean + SE),
                width = 0.1, position = position_dodge(0.9)) +
    geom_text(aes(label = round(emmean, digits = 2)), colour = "black", size = 2.5, vjust = -12,
              position = position_dodge(.9)) +
  labs(x = "Base Frequency",
       y = "Estimated Marginal Mean (N250 amplitude)",
       title = "Family Size × Base Frequency \n Interaction (N250)") +
  scale_color_custom() +
  scale_fill_custom() +
  coord_cartesian(ylim = c(-1.3, .15)) +
  theme(legend.title = element_blank())


p2 <- ggplot(emm_df, aes(x = family_size, y = emmean,
                  color = base_freq, fill = base_freq)) +
  geom_col(alpha = .4, position = position_dodge(0.9)) +
  geom_text(aes(label = round(emmean, digits = 2)), colour = "black", size = 2.5, vjust = -12,
            position = position_dodge(.9)) +
  geom_errorbar(aes(ymin = emmean - SE, ymax = emmean + SE),
                width = 0.1, position = position_dodge(0.9)) +
  labs(x = "Family Size",
       y = "Estimated Marginal Mean (N250 amplitude)",
       title = "Family Size × Base Frequency \n Interaction (N250)") +
  scale_color_custom() +
  scale_fill_custom() +
  coord_cartesian(ylim = c(-1.3, .15)) +
  theme(legend.title = element_blank())

plot_grid(p1, p2, ncol = 2)
```



3 N250 Nonword Data

3.1 Compute the ANOVA

```
anova_model_n250_nonwords <- mixed(
    value ~ Orthographic_Sensitivity * family_size * complexity +
```

```
    (1 + family_size + complexity | SubjID) +     # by-subject intercept + slopes
    (1 | SubjID:chlabel),                         # electrode nested within subject
  data   = n250_nonwords,
  method = "KR"
)
anova_model_n250_nonwords
```

```
|| Mixed Model Anova Table (Type 3 tests, KR-method)
||
|| Model: value ~ Orthographic_Sensitivity * family_size * complexity +
|| Model:     (1 + family_size + complexity | SubjID) + (1 | SubjID:chlabel)
|| Data: n250_nonwords
||                                         Effect      df      F p.value
|| 1                       Orthographic_Sensitivity   1, 59   0.05    .823
|| 2                                    family_size   1, 59   0.11    .738
|| 3                                     complexity   1, 59   0.01    .926
|| 4           Orthographic_Sensitivity:family_size   1, 59   0.00    .989
|| 5            Orthographic_Sensitivity:complexity   1, 59   0.20    .653
|| 6                        family_size:complexity 1, 1523   1.92    .166
|| 7 Orthographic_Sensitivity:family_size:complexity 1, 1523 4.58 *    .033
|| ---
|| Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '+' 0.1 ' ' 1
```

```
m1 <- anova_model_n250_nonwords$full_model    # Extract the lmer model
ranova(m1)    # Run random effects comparison
```

```
|| ANOVA-like table for random-effects: Single term deletions
||
|| Model:
|| value ~ Orthographic_Sensitivity + family_size + complexity + (1 + family_size + complexity | SubjID) + (1 | SubjID:chlabel) + Orthographic_Sen
||                                                    npar  logLik    AIC    LRT Df Pr(>Chisq)
|| <none>                                               16 -4507.1 9046.2
|| family_size in (1 + family_size + complexity | SubjID)   13 -4722.5 9471.1 430.90  3  < 2.2e-16 ***
|| complexity in (1 + family_size + complexity | SubjID)    13 -4855.6 9737.3 697.12  3  < 2.2e-16 ***
|| (1 | SubjID:chlabel)                                     15 -4708.3 9446.5 402.33  1  < 2.2e-16 ***
|| ---
|| Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Extract effect sizes from your ANOVA model
eta_squared(anova_model_n250_nonwords, partial = TRUE)
```

```
|| # Effect Size for ANOVA (Type III)
||
|| Parameter                                 | Eta2 (partial) |       95% CI
|| ------------------------------------------------------------------------------
|| Orthographic_Sensitivity                  |       8.51e-04 | [0.00, 1.00]
|| family_size                               |       1.90e-03 | [0.00, 1.00]
|| complexity                                |       1.48e-04 | [0.00, 1.00]
|| Orthographic_Sensitivity:family_size      |       2.97e-06 | [0.00, 1.00]
|| Orthographic_Sensitivity:complexity       |       3.44e-03 | [0.00, 1.00]
|| family_size:complexity                    |       1.26e-03 | [0.00, 1.00]
|| Orthographic_Sensitivity:family_size:complexity |       3.00e-03 | [0.00, 1.00]
||
|| - One-sided CIs: upper bound fixed at [1.00].
```

```
# Compute Marginal(fixed effects only) and Conditional(fixed + random effects) R²
r2(anova_model_n250_nonwords)
```

```
|| # R2 for Mixed Models
||
||   Conditional R2: 0.759
||      Marginal R2: 0.002
```

3.2 Main Effects

No main effects.

3.3 Interactions

A three way interaction between

- Sensitivity × Family Size × Complexity: significant (p =.033).

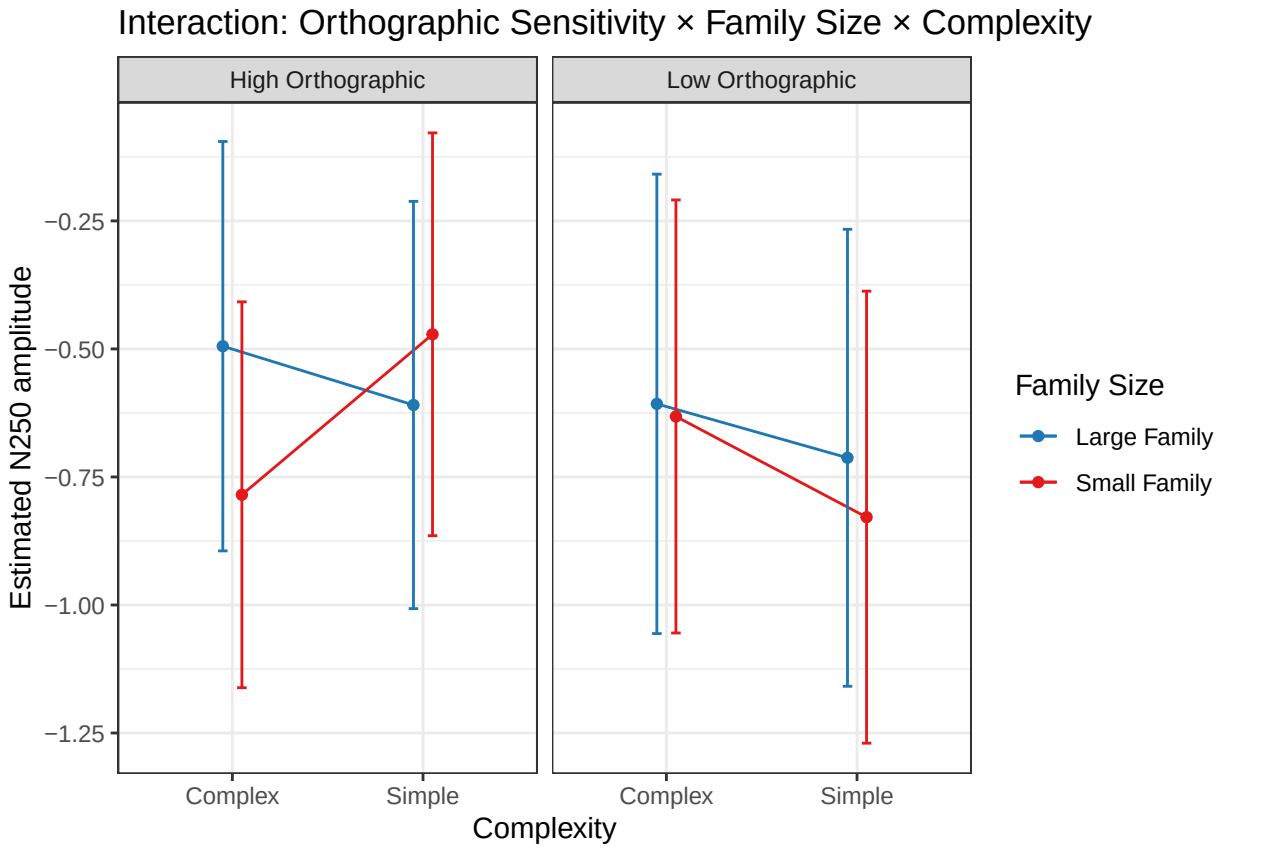
```
# Load emmeans results for the full factorial design
emm1 <- emmeans(anova_model_n250_nonwords, ~ Orthographic_Sensitivity * family_size * complexity)

# ----------- SIMPLE EFFECTS CONTRASTS -----------

# Contrast: Complexity within each Orthographic_Sensitivity × Family_Size cell
complexity_contrasts <- contrast(emm, method = "pairwise",
  by = c("Orthographic_Sensitivity", "family_size"),
```

```
  simple = "complexity", adjust = "holm")

# Contrast: Family_Size within each Orthographic_Sensitivity × Complexity cell
family_size_contrasts <- contrast(emm, method = "pairwise",
  by = c("Orthographic_Sensitivity", "complexity"),
  simple = "family_size", adjust = "holm")

# Contrast: Orthographic_Sensitivity within each Family_Size × Complexity cell
sensitivity_contrasts <- contrast(emm, method = "pairwise",
  by = c("family_size", "complexity"),
  simple = "Orthographic_Sensitivity", adjust = "holm")

# ---------- INTERACTION CONTRASTS (DIFF-OF-DIFFS) ----------

# Difference of differences: Compare all pairwise × pairwise combinations
interaction_contrasts <- contrast(emm,
  interaction = c("pairwise", "pairwise"),
  combine = TRUE, adjust = "bonferroni")

# Summarize results
summary_ic <- summary(interaction_contrasts)

# Construct human-readable contrast names
contrast_labels <- paste(
  summary_ic$Orthographic_Sensitivity_pairwise,
  summary_ic$family_size_pairwise,
  summary_ic$complexity_pairwise,
  sep = " - "
)

# ---------- COMPUTE EFFECT SIZE (COHEN'S D) ----------

# Get residual sigma from full model
sigma_val <- sigma(anova_model_n250_nonwords$full_model)

# Extract estimate, SE, and df from contrasts
est <- summary_ic$estimate
se <- summary_ic$SE
df_contr <- summary_ic$df

# Compute standardized effect size and its SE
d <- est / sigma_val
se_d <- se / sigma_val

# Compute 95% CI for d using t-distribution
tcrit <- qt(0.975, df_contr)
ci_low <- d - tcrit * se_d
ci_high <- d + tcrit * se_d

# Create table of contrasts with effect sizes
d_table <- data.frame(
  contrast = contrast_labels,
  d = d,
  se_d = se_d,
  df = df_contr,
  ci_low = ci_low,
  ci_high = ci_high
)

# View final effect size table
d_table
```

```
||                            contrast          d       se_d   df     ci_low    ci_high
|| 1  - Large Family - Small Family -  -0.5054227 0.08526265 1523 -0.6726673 -0.3381781
```

```
# 6. Plot the interaction
library(ggplot2)

emm1_df <- as.data.frame(emm1)
ggplot(emm1_df,
       aes(x = complexity, y = emmean,
           color = family_size, group = family_size)) +
  facet_wrap(~ Orthographic_Sensitivity) +
  geom_line(position = position_dodge(0.2)) +
  geom_point(position = position_dodge(0.2)) +
  geom_errorbar(aes(ymin = emmean - SE, ymax = emmean + SE),
                width = 0.1, position = position_dodge(0.2)) +
  labs(x = "Complexity", y = "Estimated N250 amplitude",
       color = "Family Size",
       title = "Interaction: Orthographic Sensitivity × Family Size × Complexity") +
  theme_bw() +
```

```
  scale_color_custom() +
  scale_fill_custom()
```



Interpretation - This is an interaction contrast (a "contrast of contrasts") across your three factors (Orthographic Sensitivity × Family Size × Complexity).

- Specifically, it is testing whether the difference (`Complex - Simple`) in `Complexity` for (`Large Family` vs. `Small Family`) differs between the two levels of `Orthographic Sensitivity`.

The contrast is asking: "Is the effect of `complexity`, in the contrast `Large` vs. `Small` family, different in `High Orthographic` vs.`Low Orthographic` participants?"

- The `estimate = 0.52` is the difference in differences (i.e. the slope difference) on your response metric (N250 amplitude).
  - $SE = 0.243$, $df = 1523$, $t = 2.140$ → yields $p = 0.0325$, so it is statistically significant (given Bonferroni correction, etc.).
- Because you used adjust = "bonferroni" and combine = TRUE, this contrast is part of a "family" of interaction contrasts that have been adjusted for multiple comparisons.

So in more conversational terms: you have evidence that High Orthographic readers show a different `complexity × family size` effect than Low Orthographic readers — in particular, in how the `complexity` effect (`Complex` vs. `Simple`) differs when comparing `Large` vs. `Small` family.

Suggests that sensitivity does influence the N250, but only in how it modulates the joint effect of family size and complexity. In other words: the way family size and complexity interact depends on whether participants are orthographically sensitive or not.

- Marginal $R^2 = 0.2\%$ → the fixed predictors (including sensitivity) account for very little variance overall.
- Conditional $R^2 = 76\%$ → most variance is indeed explained by subjects and electrodes (as anticipated).

Most of the variability in N250 amplitude reflects differences across participants and electrode sites, as expected for ERP data. Orthographic sensitivity did not produce an overall shift in N250 responses, but it did moderate the combined influence of family size and morphological complexity. This interaction was statistically significant but accounted for only a very small portion of the variance. Thus, orthographic sensitivity may play a role in how multiple lexical factors are integrated during early morphological processing, though the effect is subtle.