M21 RT Semantic Sensitivity

Joanna Morris

2025-11-10

Setup

Load libraries

1. Set `ggplot2` parameters

Load Files and Format Files

Load Files

```
#DIR <- "csv_files"
df_a <- read_csv("rt_data_hc_A.csv")
df_b <- read_csv( "rt_data_hc_B_fixed.csv")
frq_w <- read_csv("frq_cw.csv")
frq_nw <- read_csv("frq_nw.csv")
dmg <- read_csv("demo_lang_vsl_pca_hc.csv")
```

Format Files

```
# Concatenate datasets
rt <- bind_rows(AB = df_a,
                BA = df_b,
               .id = "List")
rt_dmg<- right_join(dmg, rt, join_by(SubjID == subject_nr)) |>  # Join Participant Demographic and Lang Data
                    mutate(target = tolower(target)) |>
                    filter(correct == 1)

# Divide into Experimental and Filler Items
rt_fill <- rt_dmg |> filter(str_detect(targ_type, "^FILL"))
rt_exp <- rt_dmg |> filter(!str_detect(targ_type, "^FILL"))

# Add logFS to frequency datasets
frq_w <- frq_w |> mutate(Log10FS = log10(FS))
frq_nw <- frq_nw |> mutate(Log10FS = log10(FS))

# Define Factors and Conditions
rt_exp_format <- rt_exp |>
  separate(targ_type, into = c("trial_type", "family_size", "complexity"), sep = "_",
           remove = TRUE, extra = "drop", fill = "right")

# Divide into Words and Nonwords
rt_words <- rt_exp_format |> filter(trial_type == "CW") |> select(- complexity)
rt_nwords <- rt_exp_format |> filter(trial_type == "NW")

# Join Stimulus Frequency Data
rt_words_frq <- left_join(rt_words, frq_w, join_by(target))|>
  select(-cond_trig.y, -word_trig.y) |>
  rename(cond_trig = cond_trig.x, word_trig = word_trig.x)  # remove duplicate columns
rt_nwords_frq <-  left_join(rt_nwords, frq_nw, join_by(target==word)) |>
  select(-cond_trig.y, -word_trig.y) |>
  rename(cond_trig = cond_trig.x, word_trig = word_trig.x)

# Rename BF_Split and FS_Split columns
rt_words_frq <- rt_words_frq |> rename(Base_Frequency = BF_Split, Family_Size = FS_Split) # Rename BF_Split and FS_Split columns
rt_nwords_frq <- rt_nwords_frq |> rename(Base_Frequency = BF_Split, Family_Size = FS_Split)

# Recode factor levels
# rt_words_frq <- rt_words_frq |>
#   mutate(Base_Frequency = case_match(Base_Frequency, "Low" ~ "Low BF", "High" ~ "High BF"),
```

```
#          Family_Size = case_match(Family_Size, "Small" ~ "Small Family", "Large" ~ "Large Family"))
# rt_nwords_frq <- rt_nwords_frq |> mutate(Base_Frequency = case_match(Base_Frequency, "Low" ~ "Low BF", "High" ~ "High BF"),
#                                         Family_Size = case_match(Family_Size, "Small" ~ "Small Family", "Large" ~ "Large Family"))
#
# rt_words_frq$Semantic_Sensitivity[rt_words_frq$Semantic_Sensitivity == "Low"] <- "Low Sensitivity"
# rt_words_frq$Semantic_Sensitivity[rt_words_frq$Semantic_Sensitivity == "High"] <- "High Sensitivity"
```

Word Data

Use `complete.cases()` to find which rows have missing data in the model-relevant variables:

```
# Specify only the variables used in the model
model_vars_w <- c("response_time",  "Log10BF","BF", "FS","Family_Size", "Base_Frequency", "Semantic_Sensitivity","SubjID")

# Identify incomplete rows cohort 1
incomplete_cases_words <- rt_words_frq[!complete.cases(rt_words_frq[, model_vars_w]), ]
rt_words_cmpl <- rt_words_frq[complete.cases(rt_words_frq[, model_vars_w]), ]
# View them
print(incomplete_cases_words)

# Standardize the predictors
rt_words_cmpl$Log10BF_std <- as.numeric(scale(rt_words_cmpl$Log10BF, center = TRUE, scale = TRUE))
rt_words_cmpl$FS_std <- as.numeric(scale(rt_words_cmpl$FS, center = TRUE, scale = TRUE))
rt_words_cmpl$Log10WF_std <- as.numeric(scale(rt_words_cmpl$Log10WF, center = TRUE, scale = TRUE))
rt_words_cmpl$Log10FS_std <- as.numeric(scale(rt_words_cmpl$Log10FS, center = TRUE, scale = TRUE))
rt_words_cmpl$Dim.2_std <- as.numeric(scale(rt_words_cmpl$Dim.2, center = TRUE, scale = TRUE))
```

Anova

```
anova_model_words <- mixed(
  response_time ~ Base_Frequency * Family_Size * Semantic_Sensitivity +
    (1 | SubjID) +
    (1 | STRING),
  data = rt_words_cmpl,
  method = "S")
anova_model_words
```

```
|| Mixed Model Anova Table (Type 3 tests, S-method)
||
|| Model: response_time ~ Base_Frequency * Family_Size * Semantic_Sensitivity +
|| Model:     (1 | SubjID) + (1 | STRING)
|| Data: rt_words_cmpl
||                                             Effect         df        F p.value
|| 1                                   Base_Frequency   1, 92.29 10.15 **    .002
|| 2                                      Family_Size   1, 92.30  9.28 **    .003
|| 3                             Semantic_Sensitivity   1, 64.87     0.00    .991
|| 4                       Base_Frequency:Family_Size   1, 92.29     1.01    .317
|| 5             Base_Frequency:Semantic_Sensitivity 1, 5679.51     0.41    .523
|| 6                Family_Size:Semantic_Sensitivity 1, 5679.38     0.32    .569
|| 7 Base_Frequency:Family_Size:Semantic_Sensitivity 1, 5679.34     1.03    .310
|| ---
|| Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '+' 0.1 ' ' 1
```

```
m1 <- anova_model_words$full_model    # Extract the lmer model
ranova(m1) # formally test whether adding each random effect improves fit
```

```
|| ANOVA-like table for random-effects: Single term deletions
||
|| Model:
|| response_time ~ Base_Frequency + Family_Size + Semantic_Sensitivity + (1 | SubjID) + (1 | STRING) + Base_Frequency:Family_Size + Base_Frequency
||              npar logLik   AIC     LRT Df Pr(>Chisq)
|| <none>         11 -35810 71642
|| (1 | SubjID)   10 -36768 73555 1915.39  1  < 2.2e-16 ***
|| (1 | STRING)   10 -35900 71819  179.31  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_words, partial = TRUE)
```

```
|| # Effect Size for ANOVA (Type III)
||
|| Parameter                                       | Eta2 (partial) |       95% CI
|| --------------------------------------------------------------------------------
|| Base_Frequency                                  |           0.10 | [0.02, 1.00]
|| Family_Size                                     |           0.09 | [0.02, 1.00]
|| Semantic_Sensitivity                            |       1.79e-06 | [0.00, 1.00]
|| Base_Frequency:Family_Size                      |           0.01 | [0.00, 1.00]
|| Base_Frequency:Semantic_Sensitivity             |       7.20e-05 | [0.00, 1.00]
```

```
|| Family_Size:Semantic_Sensitivity                 |        5.71e-05 | [0.00, 1.00]
|| Base_Frequency:Family_Size:Semantic_Sensitivity |        1.82e-04 | [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_words)
```

```
|| # R2 for Mixed Models
||
||   Conditional R2: 0.360
||      Marginal R2: 0.011
```

Concise Explanation

Models including random slopes for Base Frequency and Family Size by subject failed to converge or produced singular fits, indicating that the data did not support estimation of these additional variance components. Consequently, we report results from a simpler model with random intercepts for subjects and items (STRING), which converged cleanly and provided stable estimates.

Fuller explanation

We initially attempted to fit a maximal random-effects structure following Barr et al. (2013), including random slopes for Base Frequency and Family Size by subject. However, these models yielded singular fits (zero variance estimates and perfect correlations among random effects). Because such structures can produce unreliable standard errors and inflated Type I error rates, we adopted the maximal non-singular model, containing random intercepts for both subjects and items (STRING). All reported statistics are based on this model.

Brief

(A more complex model including by-subject random slopes failed to converge; results from the non-singular intercept-only model are reported.)

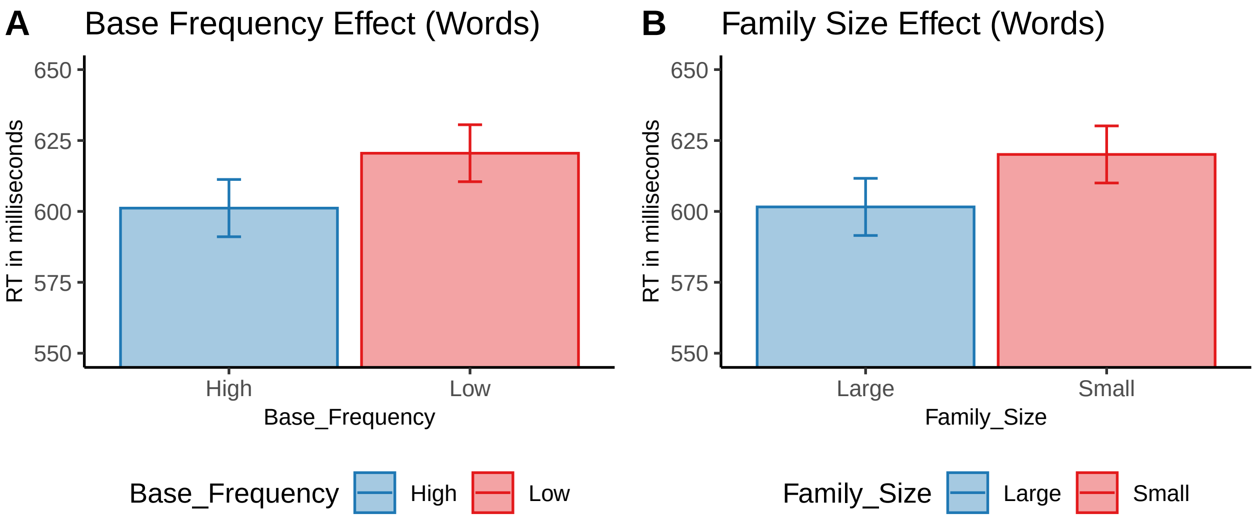
Main Findings

Effect	df	F	p.value
Base_Frequency	1, 92.29	10.15 **	.002
Family_Size	1, 92.30	9.28 **	.003

Plots

```
||  Base_Frequency   emmean       SE  df asymp.LCL asymp.UCL
||  High           601.1588 10.09536 Inf  581.3723  620.9454
||  Low            620.5137 10.04098 Inf  600.8337  640.1936
||
|| Results are averaged over the levels of: Family_Size, Semantic_Sensitivity
|| Degrees-of-freedom method: asymptotic
|| Confidence level used: 0.95
```

```
||  Family_Size   emmean       SE  df asymp.LCL asymp.UCL
||  Large       601.5859 10.07106 Inf  581.8470  621.3248
||  Small       620.0866 10.06535 Inf  600.3588  639.8143
||
|| Results are averaged over the levels of: Base_Frequency, Semantic_Sensitivity
|| Degrees-of-freedom method: asymptotic
|| Confidence level used: 0.95
```



Non-word Data

Use complete.cases() to find which rows had missing data in the model-relevant variables:

```
# Specify only the variables used in the model
# model_vars <- c("response_time", "Dim.2","SubjID")
model_vars_nw <- c("response_time","Complexity", "Family_Size", "Base_Frequency",
                   "SubjID", "ItemID", "Semantic_Sensitivity")

# Identify incomplete rows
incomplete_cases_nwords <- rt_nwords[!complete.cases(rt_nwords_frq[, model_vars_nw]), ]
rt_nwords_cmpl <- rt_nwords_frq[complete.cases(rt_nwords_frq[, model_vars_nw]), ]
# View them
print(incomplete_cases_nwords)

# str(rt_nwords_1_cmpl)
```

Standardize the predictors

```
rt_nwords_cmpl$LogBF_std <- as.numeric(scale(rt_nwords_cmpl$LogBF, center = TRUE, scale = TRUE))
rt_nwords_cmpl$FS_std <- as.numeric(scale(rt_nwords_cmpl$FS, center = TRUE, scale = TRUE))
rt_nwords_cmpl$BF_std <- as.numeric(scale(rt_nwords_cmpl$BF, center = TRUE, scale = TRUE))
rt_nwords_cmpl$Dim.2_std <- as.numeric(scale(rt_nwords_cmpl$Dim.2, center = TRUE, scale = TRUE))
```

Anova Family Size

```
rt_nwords_cmpl %>%
  count(Complexity, Base_Frequency, Semantic_Sensitivity)
```

```
|| # A tibble: 8 x 4
||   Complexity Base_Frequency Semantic_Sensitivity     n
||   <chr>      <chr>          <chr>                <int>
|| 1 Complex    High           High                   504
|| 2 Complex    High           Low                    402
|| 3 Complex    Low            High                   595
|| 4 Complex    Low            Low                    480
|| 5 Simple     High           High                   702
|| 6 Simple     High           Low                    565
|| 7 Simple     Low            High                   719
|| 8 Simple     Low            Low                    640
```

```
temp <- rt_nwords_cmpl |> filter(is.na(Complexity) & is.na(Base_Frequency))
# write_csv(temp, "temp.csv")
```

```
anova_model_nwords_fs <- mixed(
  response_time ~ Complexity * Family_Size * Semantic_Sensitivity +
    (1 | SubjID) +
    (1 | ItemID),
  data = rt_nwords_cmpl,
  method = "S")
anova_model_nwords_fs
```

```
|| Mixed Model Anova Table (Type 3 tests, S-method)
||
|| Model: response_time ~ Complexity * Family_Size * Semantic_Sensitivity +
|| Model:     (1 | SubjID) + (1 | ItemID)
|| Data: rt_nwords_cmpl
||                                      Effect         df          F p.value
|| 1                                Complexity 1, 4528.29 122.09 ***   <.001
|| 2                               Family_Size   1, 94.56       0.96    .329
|| 3                      Semantic_Sensitivity   1, 63.44       0.00    .954
|| 4                    Complexity:Family_Size 1, 4524.76       0.47    .493
|| 5           Complexity:Semantic_Sensitivity 1, 4444.26       0.16    .692
|| 6          Family_Size:Semantic_Sensitivity 1, 4442.43       0.17    .678
|| 7 Complexity:Family_Size:Semantic_Sensitivity 1, 4442.85     4.84 *    .028
|| ---
|| Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '+' 0.1 ' ' 1
```

```
m2 <- anova_model_nwords_fs$full_model    # Extract the lmer model
ranova(m2) # Run random effects comparison
```

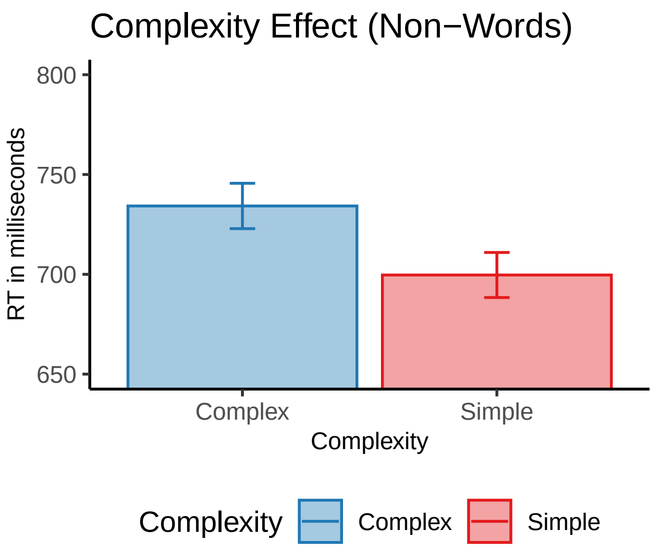
```
|| ANOVA-like table for random-effects: Single term deletions
||
|| Model:
|| response_time ~ Complexity + Family_Size + Semantic_Sensitivity + (1 | SubjID) + (1 | ItemID) + Complexity:Family_Size + Complexity:Semantic_Se
||              npar logLik   AIC     LRT Df Pr(>Chisq)
|| <none>         11 -28034 56090
|| (1 | SubjID)   10 -28903 57825 1737.56  1  < 2.2e-16 ***
|| (1 | ItemID)   10 -28105 56230  142.56  1  < 2.2e-16 ***
|| ---
|| Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Main Effects

Effect	df	F	p.value
Complexity	1, 4528.29	122.09 ***	<.001

Main Effects Plots

```
||  Complexity   emmean      SE  df asymp.LCL asymp.UCL
||  Complex    734.2425 11.3619 Inf  711.9736  756.5114
||  Simple     699.6580 11.2898 Inf  677.5304  721.7855
||
|| Results are averaged over the levels of: Family_Size, Semantic_Sensitivity
|| Degrees-of-freedom method: asymptotic
|| Confidence level used: 0.95
```



Interaction Effects

Effect	df	F	p.value
Complexity:Family__Size:Semantic__Sensitivity	1, 4442.85	4.84 *	.028

Simple Contrasts

Compare High vs Low Semantic Sensitivity within each combination of Family Size and Complexity

This gives you: 4 contrasts: one for each `Family Size × Complexity` combination. Each shows whether `High` vs `Low Semantic Sensitivity` differs significantly

If simple effects aren't significant, try looking at interaction contrasts, which test differences in the differences. You're now asking: Does the effect of Sensitivity change more in some complexity/family combinations than others?

```r
# Estimated marginal means for the family_size × complexity interaction
(emm2 <- emmeans(anova_model_nwords_fs, ~ Semantic_Sensitivity * Family_Size * Complexity))
```

```
||  Semantic_Sensitivity Family_Size Complexity emmean   SE  df asymp.LCL asymp.UCL
||  High                 Large       Complex       742 16.2 Inf       711       774
||  Low                  Large       Complex       734 16.6 Inf       702       767
||  High                 Small       Complex       726 16.2 Inf       694       758
||  Low                  Small       Complex       734 16.6 Inf       702       767
||  High                 Large       Simple        698 16.1 Inf       666       729
||  Low                  Large       Simple        706 16.4 Inf       673       738
||  High                 Small       Simple        699 16.0 Inf       668       731
||  Low                  Small       Simple        696 16.3 Inf       664       728
||
|| Degrees-of-freedom method: asymptotic
|| Confidence level used: 0.95
```

```r
# Get all pairwise contrasts
emm2_contrasts <- contrast(emm2,  method = "pairwise", by = NULL, adjust = "none")
# emm2_contrasts

# Keep only the contrasts you want
# Simple effects of family_size at each level of complexity
```

```r
# Simple effects of complexity at each level of family_size
keep2 <- c("High Large Complex - High Large Simple",
           "High Small Complex - High Small Simple",
           "Low Large Complex - Low Large Simple",
           "Low Small Complex - Low Small Simple",
           "High Large Complex - High Small Complex",
           "High Large Simple - High Small Simple",
           "Low Large Complex - Low Small Complex",
           "Low Large Simple - Low Small Simple",
           "High Large Complex - Low Large Complex",
           "High Small Complex - Low Small Complex",
           "High Large Simple - Low Small Simple",
           "High Small Simple - Low Small Simple")

(emm2_contrasts_filtered <- subset(emm2_contrasts, contrast %in% keep2))
```

```
||  contrast                                estimate    SE  df z.ratio p.value
||  High Large Complex - Low Large Complex     7.974 22.60 Inf   0.353  0.7239
||  High Large Complex - High Small Complex   16.270  8.25 Inf   1.973  0.0485
||  High Large Complex - High Large Simple    44.698  5.97 Inf   7.484  <.0001
||  Low Large Complex - Low Small Complex      0.216  8.83 Inf   0.024  0.9805
||  Low Large Complex - Low Large Simple      28.755  6.63 Inf   4.339  <.0001
||  High Small Complex - Low Small Complex    -8.080 22.50 Inf  -0.359  0.7197
||  High Small Complex - High Small Simple    26.910  5.80 Inf   4.641  <.0001
||  Low Small Complex - Low Small Simple      37.975  6.37 Inf   5.966  <.0001
||  High Large Simple - High Small Simple     -1.519  7.67 Inf  -0.198  0.8430
||  High Large Simple - Low Small Simple       1.466 22.90 Inf   0.064  0.9490
||  Low Large Simple - Low Small Simple        9.435  8.01 Inf   1.177  0.2391
||  High Small Simple - Low Small Simple       2.985 22.30 Inf   0.134  0.8933
||
|| Degrees-of-freedom method: asymptotic
```

```r
# Get Confidence Intervals
(emm2_contrasts_filtered_ci <- confint(emm2_contrasts_filtered))
```

```
||  contrast                                estimate    SE  df asymp.LCL asymp.UCL
||  High Large Complex - Low Large Complex     7.974 22.60 Inf   -36.261      52.2
||  High Large Complex - High Small Complex   16.270  8.25 Inf     0.106      32.4
||  High Large Complex - High Large Simple    44.698  5.97 Inf    32.992      56.4
||  Low Large Complex - Low Small Complex      0.216  8.83 Inf   -17.081      17.5
||  Low Large Complex - Low Large Simple      28.755  6.63 Inf    15.766      41.7
||  High Small Complex - Low Small Complex    -8.080 22.50 Inf   -52.202      36.0
||  High Small Complex - High Small Simple    26.910  5.80 Inf    15.545      38.3
||  Low Small Complex - Low Small Simple      37.975  6.37 Inf    25.499      50.5
||  High Large Simple - High Small Simple     -1.519  7.67 Inf   -16.550      13.5
||  High Large Simple - Low Small Simple       1.466 22.90 Inf   -43.475      46.4
||  Low Large Simple - Low Small Simple        9.435  8.01 Inf    -6.273      25.1
||  High Small Simple - Low Small Simple       2.985 22.30 Inf   -40.637      46.6
||
|| Degrees-of-freedom method: asymptotic
|| Confidence level used: 0.95
```

```r
# Get effect sizes
# Get all pairwise effect sizes
effs2 <- eff_size(emm2, sigma = sigma(m2), edf = df.residual(m2))

# Remove the  redundant rows
(effs2_filtered <- subset(effs2, contrast %in% keep2))
```

```
||  contrast                                effect.size     SE  df asymp.LCL asymp.UCL
||  High Large Complex - Low Large Complex      0.07804 0.2210 Inf  -0.35489     0.511
||  High Large Complex - High Small Complex     0.15923 0.0807 Inf   0.00101     0.317
||  High Large Complex - High Large Simple      0.43747 0.0586 Inf   0.32255     0.552
||  Low Large Complex - Low Small Complex       0.00211 0.0864 Inf  -0.16717     0.171
||  Low Large Complex - Low Large Simple        0.28143 0.0649 Inf   0.15417     0.409
||  High Small Complex - Low Small Complex     -0.07908 0.2200 Inf  -0.51091     0.353
||  High Small Complex - High Small Simple      0.26337 0.0568 Inf   0.15201     0.375
||  Low Small Complex - Low Small Simple        0.37166 0.0624 Inf   0.24933     0.494
||  High Large Simple - High Small Simple      -0.01486 0.0751 Inf  -0.16197     0.132
||  High Large Simple - Low Small Simple        0.01435 0.2240 Inf  -0.42550     0.454
||  Low Large Simple - Low Small Simple         0.09234 0.0784 Inf  -0.06140     0.246
||  High Small Simple - Low Small Simple        0.02922 0.2180 Inf  -0.39772     0.456
||
|| sigma used for effect sizes: 102.2
|| Degrees-of-freedom method: inherited from asymptotic when re-gridding
|| Confidence level used: 0.95
```

Interaction Contrasts The interaction contrast tests whether the difference in the complexity effect for large vs small families differs across sensitivity?

$$[[(A_1 - A_2) \text{ in } B_1] - [(A_1 - A_2) \text{ in } B_2] \text{ in Condition } C_1] - [[(A_1 - A_2) in B_1] - [(A_1 - A_2) in B_2] \text{in Condition } C_2]$$

```r
#  Interaction contrasts (difference-of-differences)
#    Compare complexity effect in large vs small family)
contrast(emm2, interaction = "pairwise", by = NULL, adjust = "holm")
```

```
||  Semantic_Sensitivity_pairwise Family_Size_pairwise Complexity_pairwise estimate   SE  df z.ratio p.value
||  High - Low                    Large - Small        Complex - Simple          27 12.3 Inf   2.199  0.0279
||
|| Degrees-of-freedom method: asymptotic
```

```r
confint(contrast(emm2, interaction = c("pairwise", "pairwise")))
```

```
||  Semantic_Sensitivity_pairwise Family_Size_pairwise Complexity_pairwise estimate   SE  df asymp.LCL asymp.UCL
||  High - Low                    Large - Small        Complex - Simple          27 12.3 Inf      2.94      51.1
||
|| Degrees-of-freedom method: asymptotic
|| Confidence level used: 0.95
```

```r
# Compute the A1 - A2 difference within each combination of B × C
(complexity_diff <- contrast(emm2, method = "revpairwise",
                            by = c("Semantic_Sensitivity", "Family_Size"),
                            simple = "Complexity"))
```

```
|| Semantic_Sensitivity = High, Family_Size = Large:
||  contrast         estimate   SE  df z.ratio p.value
||  Simple - Complex    -44.7 5.97 Inf  -7.484  <.0001
||
|| Semantic_Sensitivity = Low, Family_Size = Large:
||  contrast         estimate   SE  df z.ratio p.value
||  Simple - Complex    -28.8 6.63 Inf  -4.339  <.0001
||
|| Semantic_Sensitivity = High, Family_Size = Small:
||  contrast         estimate   SE  df z.ratio p.value
||  Simple - Complex    -26.9 5.80 Inf  -4.641  <.0001
||
|| Semantic_Sensitivity = Low, Family_Size = Small:
||  contrast         estimate   SE  df z.ratio p.value
||  Simple - Complex    -38.0 6.37 Inf  -5.966  <.0001
||
|| Degrees-of-freedom method: asymptotic
```

```r
# Compute how that A-effect changes across the levels of B, separately for each level of C
(family_size_complexity_int_within_sensitivity <- contrast(complexity_diff,
                                                          method = "revpairwise",
                                                          by = "Semantic_Sensitivity", simple = "Family_Size"))
```

```
|| contrast = Simple - Complex, Semantic_Sensitivity = High:
||  contrast1     estimate   SE  df z.ratio p.value
||  Small - Large    17.79 8.32 Inf   2.139  0.0325
||
|| contrast = Simple - Complex, Semantic_Sensitivity = Low:
||  contrast1     estimate   SE  df z.ratio p.value
||  Small - Large    -9.22 9.18 Inf  -1.004  0.3152
||
|| Degrees-of-freedom method: asymptotic
```

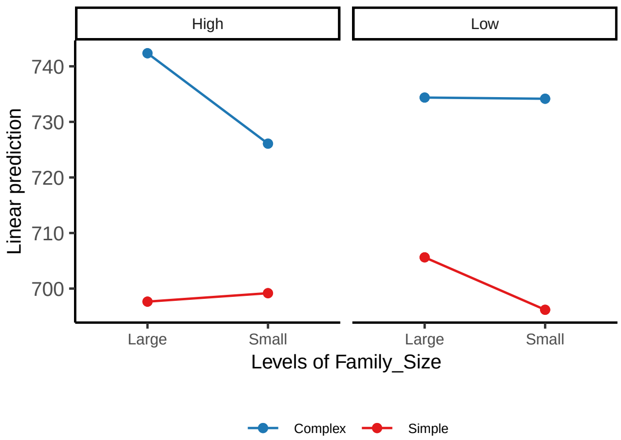
```r
# Get confidence intervals
confint(family_size_complexity_int_within_sensitivity)
```

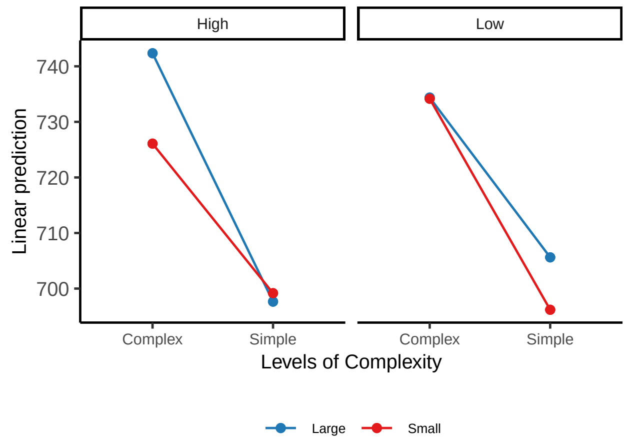
```
|| contrast = Simple - Complex, Semantic_Sensitivity = High:
||  contrast1     estimate   SE  df asymp.LCL asymp.UCL
||  Small - Large    17.79 8.32 Inf      1.49     34.09
||
|| contrast = Simple - Complex, Semantic_Sensitivity = Low:
||  contrast1     estimate   SE  df asymp.LCL asymp.UCL
||  Small - Large    -9.22 9.18 Inf    -27.21      8.77
||
|| Degrees-of-freedom method: asymptotic
|| Confidence level used: 0.95
```

Interaction Plots

```r
p4 <- emmip(anova_model_nwords_fs, Complexity ~ Family_Size | Semantic_Sensitivity) + my_style
p5 <- emmip(anova_model_nwords_fs, Family_Size ~ Complexity | Semantic_Sensitivity) + my_style

plot_grid(p4, p5, ncol = 2)
```





Responses were slower to *Complex* than *Simple* nonwords in every condition. The *Complexity effect* (`Complex - Simple`) varied with both *Family Size* and *Semantic Sensitivity*.

The Complexity effect (slower responses for complex vs. simple nonwords) is robust across all groups.

However, its magnitude varies:

- Among **high-sensitivity participants**, the effect is larger for large families ($\approx$ 45 ms) than small families ($\approx$ 27 ms).
- Among **low-sensitivity participants**, the pattern reverses slightly ($\approx$ 29 ms vs. 38 ms).

The difference in the Complexity $\times$ Family Size interaction between high- and low-sensitivity participants is about 27 ms.

- **High-sensitivity participants** showed a stronger complexity effect for large-family nonwords than for small-family ones.
- **Low-sensitivity participants** showed the opposite or no difference.

This indicates that semantic sensitivity modulates how morphological family size influences the cost of morphological complexity in nonword processing.

All groups show reliable complexity effects (complex slower than simple). Only one cross-condition difference is significant: High-sensitivity participants respond faster to complex nonwords from small families than to complex nonwords from large-families.

Summary interpretation (for Results section):

Response times to morphologically complex nonwords were significantly slower than to simple nonwords, indicating greater processing cost for complexity. While overall family size and semantic sensitivity did not produce main effects, there was a significant `Complexity × Family Size × Semantic Sensitivity` interaction ($p = .028$).

Follow-up contrasts showed that for participants with high semantic sensitivity, the complexity effect was larger for large-family nonwords ($\approx$ 45 ms) than for small-family nonwords ($\approx$ 27 ms). In contrast, participants with low semantic sensitivity showed little difference or the reverse pattern. This suggests that individuals with greater semantic knowledge are more sensitive to morphological family size cues when processing novel morphological structures, showing amplified complexity costs when nonwords resemble rich morphological families.

Anova Base Frequency

```r
anova_model_nwords_bf <- mixed(
  response_time ~ Complexity * Base_Frequency * Semantic_Sensitivity +
    (1 | SubjID) +
    (1 | ItemID),
  data = rt_nwords_cmpl,
  method = "S")
anova_model_nwords_bf
```

```
|| Mixed Model Anova Table (Type 3 tests, S-method)
||
|| Model: response_time ~ Complexity * Base_Frequency * Semantic_Sensitivity +
|| Model:     (1 | SubjID) + (1 | ItemID)
|| Data: rt_nwords_cmpl
||                                        Effect         df          F p.value
|| 1                                  Complexity 1, 4533.26 125.15 ***   <.001
|| 2                              Base_Frequency   1, 95.24  12.70 ***   <.001
|| 3                        Semantic_Sensitivity   1, 63.45       0.00    .968
|| 4                   Complexity:Base_Frequency 1, 4535.00     3.92 *    .048
|| 5             Complexity:Semantic_Sensitivity 1, 4446.84       0.21    .647
|| 6         Base_Frequency:Semantic_Sensitivity 1, 4445.63       1.15    .284
|| 7 Complexity:Base_Frequency:Semantic_Sensitivity 1, 4446.76    2.56    .110
|| ---
|| Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '+' 0.1 ' ' 1
```

```r
m3 <- anova_model_nwords_bf$full_model    # Extract the lmer model
ranova(m3) # Run random effects comparison
```

```
|| ANOVA-like table for random-effects: Single term deletions
||
|| Model:
|| response_time ~ Complexity + Base_Frequency + Semantic_Sensitivity + (1 | SubjID) + (1 | ItemID) + Complexity:Base_Frequency + Complexity:Seman
||              npar logLik   AIC     LRT Df Pr(>Chisq)
|| <none>         11 -28028 56077
|| (1 | SubjID)   10 -28897 57815 1739.59  1  < 2.2e-16 ***
|| (1 | ItemID)   10 -28089 56197  122.45  1  < 2.2e-16 ***
|| ---
|| Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Main Findings

Effect	df	F	p.value
Complexity	1, 4533.26	125.15 ***	<.001
Base_Frequency	1, 95.24	12.70 **	<.001
Complexity:Base_Frequency	1, 4535.47	3.92 *	.048

Participants responded more slowly to complex nonwords and to low-frequency-base nonwords.

Interaction Effects: `Complexity x Base_Frequency`

```r
# Estimated marginal means for the family_size × base frequency interaction
(emm1 <- emmeans(anova_model_nwords_bf, ~ Complexity * Base_Frequency))
```

Simple Contrasts

```
||  Complexity Base_Frequency emmean   SE  df asymp.LCL asymp.UCL
||  Complex    High              748 11.9 Inf       725       772
||  Simple     High              707 11.7 Inf       684       730
||  Complex    Low               721 11.8 Inf       698       744
||  Simple     Low               692 11.7 Inf       669       715
||
|| Results are averaged over the levels of: Semantic_Sensitivity
|| Degrees-of-freedom method: asymptotic
|| Confidence level used: 0.95
```

```r
# Get all pairswise contrasts
emm1_contrasts <- contrast(emm1,  method = "pairwise", by = NULL, adjust = "none")
emm1_contrasts
```

```
||  contrast                    estimate   SE  df z.ratio p.value
||  Complex High - Simple High      41.3 4.63 Inf   8.915  <.0001
||  Complex High - Complex Low      27.3 6.93 Inf   3.945  0.0001
||  Complex High - Simple Low       56.2 6.76 Inf   8.319  <.0001
||  Simple High - Complex Low      -14.0 6.66 Inf  -2.099  0.0359
||  Simple High - Simple Low        14.9 6.48 Inf   2.298  0.0216
||  Complex Low - Simple Low        28.9 4.24 Inf   6.800  <.0001
```

```
||
|| Results are averaged over the levels of: Semantic_Sensitivity
|| Degrees-of-freedom method: asymptotic
```

```r
# Keep only the contrasts you want
# Simple effects of Complexity at each level of Base_Frequency
# Simple effects of Base_Frequency at each level of Complexity
keep <- c("Complex High - Simple High",
          "Complex Low - Simple Low",
          "Complex High - Complex Low",
          "Simple High - Simple Low")
(emm1_contrasts_filtered <- subset(emm1_contrasts, contrast %in% keep))
```

```
||  contrast                   estimate   SE  df z.ratio p.value
||  Complex High - Simple High     41.3 4.63 Inf   8.915  <.0001
||  Complex High - Complex Low     27.3 6.93 Inf   3.945  0.0001
||  Simple High - Simple Low       14.9 6.48 Inf   2.298  0.0216
||  Complex Low - Simple Low       28.9 4.24 Inf   6.800  <.0001
||
|| Results are averaged over the levels of: Semantic_Sensitivity
|| Degrees-of-freedom method: asymptotic
```

```r
# Get Confidence Intervals
(emm1_contrasts_filtered_ci <- confint(emm1_contrasts_filtered))
```

```
||  contrast                   estimate   SE  df asymp.LCL asymp.UCL
||  Complex High - Simple High     41.3 4.63 Inf     32.23      50.4
||  Complex High - Complex Low     27.3 6.93 Inf     13.76      40.9
||  Simple High - Simple Low       14.9 6.48 Inf      2.19      27.6
||  Complex Low - Simple Low       28.9 4.24 Inf     20.54      37.2
||
|| Results are averaged over the levels of: Semantic_Sensitivity
|| Degrees-of-freedom method: asymptotic
|| Confidence level used: 0.95
```

```r
# Get effect sizes
# Get all pairwise effect sizes
effs1 <- eff_size(emm1, sigma = sigma(m3), edf = df.residual(m3))
effs1
```

```
||  contrast                   effect.size     SE  df asymp.LCL asymp.UCL
||  Complex High - Simple High       0.405 0.0456 Inf    0.3152     0.494
||  Complex High - Complex Low       0.268 0.0679 Inf    0.1346     0.401
||  Complex High - Simple Low        0.550 0.0664 Inf    0.4202     0.681
||  Simple High - Complex Low       -0.137 0.0652 Inf   -0.2645    -0.009
||  Simple High - Simple Low         0.146 0.0635 Inf    0.0214     0.270
||  Complex Low - Simple Low         0.283 0.0417 Inf    0.2009     0.364
||
|| Results are averaged over the levels of: Semantic_Sensitivity
|| sigma used for effect sizes: 102.1
|| Degrees-of-freedom method: inherited from asymptotic when re-gridding
|| Confidence level used: 0.95
```

```r
# Remove the two redundant rows (rows 3 and 4)
(effs1_filtered <- subset(effs1, !contrast %in% c("Complex High - Simple Low",
                                                  "Simple High - Complex Low")))
```

```
||  contrast                   effect.size     SE  df asymp.LCL asymp.UCL
||  Complex High - Simple High       0.405 0.0456 Inf    0.3152     0.494
||  Complex High - Complex Low       0.268 0.0679 Inf    0.1346     0.401
||  Simple High - Simple Low         0.146 0.0635 Inf    0.0214     0.270
||  Complex Low - Simple Low         0.283 0.0417 Inf    0.2009     0.364
||
|| Results are averaged over the levels of: Semantic_Sensitivity
|| sigma used for effect sizes: 102.1
|| Degrees-of-freedom method: inherited from asymptotic when re-gridding
|| Confidence level used: 0.95
```

```r

#  Interaction contrasts (difference-of-differences)
#    Compare base frequency effect in large vs small family)
contrast(emm1, interaction = "pairwise", by = NULL, adjust = "holm")
```

Interaction Contrasts

```
||  Complexity_pairwise Base_Frequency_pairwise estimate  SE  df z.ratio p.value
||  Complex - Simple    High - Low                  12.5 6.3 Inf   1.979  0.0479
||
|| Results are averaged over the levels of: Semantic_Sensitivity
|| Degrees-of-freedom method: asymptotic
```

```r
# Get confidence intervals, for each base frequency effect for each family size and then for interaction effect
confint(contrast(emmeans(m3, ~ Complexity | Base_Frequency), "pairwise"))
```

```
|| Base_Frequency = High:
||  contrast         estimate   SE  df asymp.LCL asymp.UCL
||  Complex - Simple     41.3 4.63 Inf      32.2      50.4
||
|| Base_Frequency = Low:
||  contrast         estimate   SE  df asymp.LCL asymp.UCL
||  Complex - Simple     28.9 4.24 Inf      20.5      37.2
||
|| Results are averaged over the levels of: Semantic_Sensitivity
|| Degrees-of-freedom method: asymptotic
|| Confidence level used: 0.95
```

```
confint(contrast(emm1, interaction = c("pairwise", "pairwise")))
```

```
||  Complexity_pairwise Base_Frequency_pairwise estimate  SE  df asymp.LCL asymp.UCL
||  Complex - Simple    High - Low                  12.5 6.3 Inf     0.118      24.8
||
|| Results are averaged over the levels of: Semantic_Sensitivity
|| Degrees-of-freedom method: asymptotic
|| Confidence level used: 0.95
```

A small but reliable Complexity × Base Frequency interaction ($p = .048$) suggests that the complexity cost was smaller for nonwords derived from low-frequency bases.

Complexity	Base Frequency	Mean RT (ms)	Interpretation
Complex	High	748	slowest
Simple	High	707	41 ms faster
Complex	Low	721	28 ms slower than Simple Low
Simple	Low	692	fastest

Both complexity and base frequency affect RTs additively, but their combination reveals that high-frequency bases magnify the complexity cost.

- The complexity effect (`Complex - Simple`) is larger for *high-frequency* bases (41 ms) than for *low-frequency* ones (29 ms).
- The base-frequency advantage (`High - Low`) is stronger for *complex* items (27 ms) than for *simple* ones (15 ms).
- Both effects are moderate in size (Cohen's $d \approx 0.3$ˇ0.4).

The complexity cost increases by about 12 ms when the base is high frequency rather than low frequency, confirming the small but significant interaction.

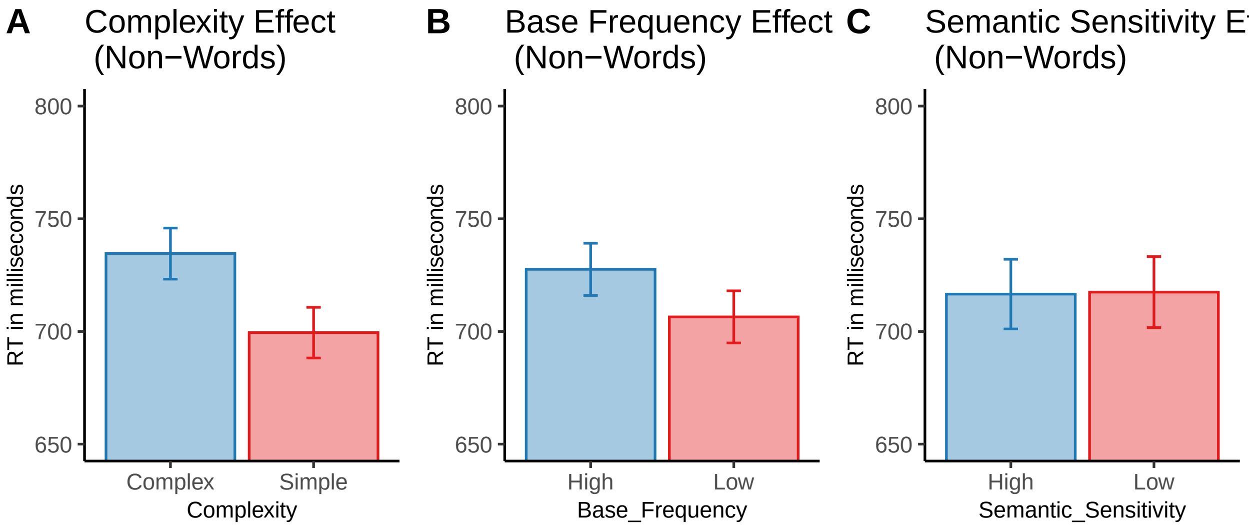
No effects involving Semantic Sensitivity were observed, indicating that this base-frequency modulation of complexity applies broadly across participants, independent of their semantic knowledge.

Main Effects Plots

```
||  Complexity   emmean       SE  df asymp.LCL asymp.UCL
||  Complex    734.5516 11.32891 Inf  712.3473  756.7559
||  Simple     699.4623 11.25555 Inf  677.4018  721.5227
||
|| Results are averaged over the levels of: Base_Frequency, Semantic_Sensitivity
|| Degrees-of-freedom method: asymptotic
|| Confidence level used: 0.95
```

```
||  Base_Frequency   emmean       SE  df asymp.LCL asymp.UCL
||  High           727.5677 11.57476 Inf  704.8815  750.2538
||  Low            706.4462 11.56316 Inf  683.7828  729.1096
||
|| Results are averaged over the levels of: Complexity, Semantic_Sensitivity
|| Degrees-of-freedom method: asymptotic
|| Confidence level used: 0.95
```

```
||  Semantic_Sensitivity   emmean       SE  df asymp.LCL asymp.UCL
||  High                 716.5652 15.47076 Inf  686.2431  746.8874
||  Low                  717.4486 15.74909 Inf  686.5810  748.3163
||
|| Results are averaged over the levels of: Complexity, Base_Frequency
|| Degrees-of-freedom method: asymptotic
|| Confidence level used: 0.95
```



Interaction Plots

```r
p9 <- emmip(anova_model_nwords_bf, Complexity ~ Base_Frequency) + my_style
p10 <- emmip(anova_model_nwords_bf, Base_Frequency ~ Complexity) + my_style

plot_grid(p9, p10, ncol = 2)
```