Analysis

Zhaosen Guo

7/10/2021

Intro

For the given psycholinguistics experiment data, I have generated this R Markdown file for easier access to the data, codes, and explanations. This file can also generate reports in html/pdf when pressing the "Knit" option in RStudio.
(FYI, a R Markdown cheat sheet is included in the folder.)

Here are the required libraries:

```
library(tidyverse)
library(lme4)
library(lmerTest)
```

We will be running the following mixed-effect models with variables and parameters -

Language	Independent Variable (Fixed effect)	Dependent Variable	Random Effects	Link
English	sentence context (scope & negation)	sentence length (ms)	subjects & keywords	jump
Japanese	sentence context (scope & negation)	sentence length (ms)	subjects & keywords	jump
English	sentence context (scope & negation)	fundamental frequency (Hz)	subjects & keywords	jump
Japanese	sentence context (scope & negation)	fundamental frequency (Hz)	subjects & keywords	jump
English	sentence context (scope ##*w/out "Wool"* & negation)	sentence length (ms)	subjects & keywords	jump
English	sentence context (scope ##*w/out "Wool"* & negation)	fundamental frequency (Hz)	subjects & keywords	jump

Models

Context vs SentenceLength, English

back to top

First, import the data set.

```
data.SenLen.Eng <- read_csv("data/SentenceLengthEnglish.csv")
```

The code below shows that the class types of the data are not correct.

```
sapply(data.SenLen.Eng, class)
```

```
       subject        sentence         context sentence_length           maxF0
   "character"     "character"     "character"       "numeric"       "numeric"
        meanF0              F1              F2              F3         keyword
     "numeric"       "numeric"       "numeric"       "numeric"     "character"
 last_word_len
     "numeric"
```

Change them from character class to factor class.

```
col_target <- c("subject", "sentence", "context", "keyword")
data.SenLen.Eng[col_target] <- lapply(data.SenLen.Eng[col_target], as.factor)
```

Now the mixed-effect model:

```
model.SenLen.Eng<- lmer(data = data.SenLen.Eng,
                        sentence_length ~ context +
                          (1 + context||keyword) +
                          (1 + context||subject), REML = FALSE,
                        control=lmerControl(check.conv.singular = .makeCC(action = "ignore",  tol = 1e-
summary(model.SenLen.Eng)
```

```
Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
  method [lmerModLmerTest]
Formula:
sentence_length ~ context + (1 + context || keyword) + (1 + context ||
    subject)
   Data: data.SenLen.Eng
Control: lmerControl(check.conv.singular = .makeCC(action = "ignore",
    tol = 1e-04))

     AIC      BIC   logLik deviance df.resid
  3741.2   3836.3  -1843.6   3687.2      223

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.9073 -0.3670 -0.0913  0.2323 10.2280

Random effects:
 Groups    Name                    Variance  Std.Dev.  Corr
 subject   contextall > not        3.980e+04 199.51107
           contextmost/many > not  3.049e+04 174.61672 0.99
           contextnot > all        1.111e+05 333.24640 0.86 0.79
           contextnot > most/many  3.147e+04 177.40585 0.92 0.96 0.58
 subject.1 (Intercept)             2.807e+00   1.67548
 keyword   contextall > not        2.449e+04 156.47752
```

```
           contextmost/many > not 2.857e+04 169.03330 0.67
           contextnot > all       5.364e+03  73.24233 1.00 0.67
           contextnot > most/many 2.813e+04 167.71720 0.67 1.00 0.67
 keyword.1 (Intercept)            3.396e-03   0.05827
 Residual                         1.154e+05 339.68167
Number of obs: 250, groups:  subject, 25; keyword, 5

Fixed effects:
                       Estimate Std. Error       df t value Pr(>|t|)
(Intercept)            1964.541    106.266    4.013  18.487 4.92e-05 ***
contextmost/many > not   36.135    162.287    4.142   0.223    0.834
contextnot > all         29.971     82.819    6.112   0.362    0.730
contextnot > most/many  107.914    162.246    4.224   0.665    0.541
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) cnt/>n cntx>a
cntxtmst/>n -0.575
cntxtnt>all -0.590  0.421
cntxtn>mst/ -0.580  0.910  0.392
```

From the last section of the "Fixed effects" above, we can see that non of the context level displayed statistical significance in influencing the sentence length.

Context vs SentenceLength, Japanese

back to top

Import and process the data:

```
data.SenLen.Jap <- read_csv("data/SentenceLengthJapanese.csv")
data.SenLen.Jap[col_target] <- lapply(data.SenLen.Jap[col_target], as.factor)
```

Model:

```
model.SenLen.Jap <- lmer(data = data.SenLen.Jap,
                         sentence_length ~ context +
                           (1 + context||keyword) +
                           (1 + context||subject), REML = FALSE,
                         control=lmerControl(check.conv.singular = .makeCC(action = "ignore",  tol = 1e-
```

```
Warning: Model failed to converge with 1 negative eigenvalue: -1.4e-02
```

```
summary(model.SenLen.Jap)
```

```
Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
  method [lmerModLmerTest]
Formula:
sentence_length ~ context + (1 + context || keyword) + (1 + context ||
    subject)
   Data: data.SenLen.Jap
```

```
Control: lmerControl(check.conv.singular = .makeCC(action = "ignore",
    tol = 1e-04))

     AIC      BIC   logLik deviance df.resid
  2894.0   2983.1  -1420.0   2840.0      173

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.5317 -0.5746 -0.1197  0.5512  4.5462

Random effects:
 Groups    Name                    Variance  Std.Dev.  Corr
 subject   contextall > not        9.441e+04 307.26019
           contextmost/many > not  1.683e+05 410.30193 1.00
           contextnot > all        9.691e+04 311.31009 0.99  0.99
           contextnot > most/many  1.007e+05 317.29146 0.93  0.95  0.89
 subject.1 (Intercept)             3.049e+00   1.74600
 keyword   contextall > not        1.266e+05 355.87074
           contextmost/many > not  1.618e+04 127.20175 -0.30
           contextnot > all        1.110e+05 333.10654  1.00 -0.23
           contextnot > most/many  2.655e+04 162.95350 -0.30  1.00 -0.23
 keyword.1 (Intercept)             5.501e-04   0.02345
 Residual                          5.622e+04 237.10412
Number of obs: 200, groups:  subject, 20; keyword, 5

Fixed effects:
                        Estimate Std. Error        df t value Pr(>|t|)
(Intercept)            2825.8693   218.7972    3.6743  12.915  0.00034 ***
contextmost/many > not   22.9642   230.6549    4.2104   0.100  0.92525
contextnot > all         29.8974    48.5664    3.0828   0.616  0.58063
contextnot > most/many    0.2576   241.8675    4.6671   0.001  0.99919
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) cnt/>n cntx>a
cntxtmst/>n -0.824
cntxtnt>all -0.351  0.334
cntxtn>mst/ -0.819  0.961  0.305
```

Similar to the English result, there's no significant correlation between sentence context and sentence length in Japanese.

Context vs Fundamental Frequency (F0), English

back to top

```
model.meanF0.Eng <- lmer(data = data.SenLen.Eng,
                         meanF0 ~ context +
                           (1 + context||keyword) +
```

```
                          (1 + context||subject), REML = FALSE,
                         control=lmerControl(check.conv.singular = .makeCC(action = "ignore",  tol = 1e-
summary(model.meanF0.Eng)
```

Model for the mean fundamental frequency during the sentence (meanF0):

```
Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
  method [lmerModLmerTest]
Formula: meanF0 ~ context + (1 + context || keyword) + (1 + context ||
    subject)
   Data: data.SenLen.Eng
Control: lmerControl(check.conv.singular = .makeCC(action = "ignore",
    tol = 1e-04))

     AIC      BIC   logLik deviance df.resid
  2221.6   2316.6  -1083.8   2167.6      223

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.5773 -0.5823 -0.0529  0.4831  3.4514

Random effects:
 Groups    Name                    Variance  Std.Dev.  Corr
 subject   contextall > not        1.760e+03 4.195e+01
           contextmost/many > not  1.649e+03 4.061e+01 0.98
           contextnot > all        1.861e+03 4.314e+01 1.00  0.99
           contextnot > most/many  2.035e+03 4.511e+01 0.99  1.00  1.00
 subject.1 (Intercept)             7.697e-03 8.773e-02
 keyword   contextall > not        5.402e+01 7.349e+00
           contextmost/many > not  4.308e-07 6.564e-04 -1.00
           contextnot > all        9.769e+00 3.125e+00  1.00 -1.00
           contextnot > most/many  1.563e-06 1.250e-03 -0.53  0.53 -0.53
 keyword.1 (Intercept)             0.000e+00 0.000e+00
 Residual                          2.038e+02 1.428e+01
Number of obs: 250, groups:  subject, 25; keyword, 5

Fixed effects:
                      Estimate Std. Error      df t value Pr(>|t|)
(Intercept)            168.773      9.546  23.201  17.681 5.83e-15 ***
contextmost/many > not   3.515      5.307   5.351   0.662    0.535
contextnot > all        -4.913      3.438   4.074  -1.429    0.225
contextnot > most/many   6.879      5.155   4.835   1.334    0.241
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) cnt/>n cntx>a
cntxtmst/>n -0.486
cntxtnt>all -0.343  0.772
cntxtn>mst/ -0.327  0.827  0.787
```

```
model.maxF0.Eng <- lmer(data = data.SenLen.Eng,
                        maxF0 ~ context +
                          (1 + context||keyword) +
                          (1 + context||subject), REML = FALSE,
                        control=lmerControl(check.conv.singular = .makeCC(action = "ignore",  tol = 1e-
```

Model for the max fundamental frequency during the sentence (maxF0):

```
Warning: Model failed to converge with 1 negative eigenvalue: -7.6e-03
```

```
summary(model.maxF0.Eng)
```

```
Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
  method [lmerModLmerTest]
Formula: maxF0 ~ context + (1 + context || keyword) + (1 + context ||
    subject)
   Data: data.SenLen.Eng
Control: lmerControl(check.conv.singular = .makeCC(action = "ignore",
    tol = 1e-04))

     AIC      BIC   logLik deviance df.resid
  2761.7   2856.8  -1353.9   2707.7      223

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.0396 -0.6609 -0.1358  0.6050  2.6685

Random effects:
 Groups    Name                    Variance  Std.Dev.  Corr
 subject   contextall > not        3.343e+03 5.782e+01
           contextmost/many > not  3.375e+03 5.810e+01 0.77
           contextnot > all        3.138e+03 5.601e+01 1.00 0.82
           contextnot > most/many  4.378e+03 6.616e+01 0.90 0.97 0.93
 subject.1 (Intercept)             1.057e-03 3.252e-02
 keyword   contextall > not        0.000e+00 0.000e+00
           contextmost/many > not  2.266e+01 4.760e+00  NaN
           contextnot > all        5.223e-07 7.227e-04  NaN 0.54
           contextnot > most/many  1.114e+01 3.338e+00  NaN 1.00 0.54
 keyword.1 (Intercept)             0.000e+00 0.000e+00
 Residual                          2.060e+03 4.538e+01
Number of obs: 250, groups:  subject, 25; keyword, 5

Fixed effects:
                       Estimate Std. Error      df t value Pr(>|t|)
(Intercept)             263.061     12.695  25.294  20.721   <2e-16 ***
contextmost/many > not  -11.968     11.923  11.262  -1.004    0.337
contextnot > all         -4.653      7.482 148.169  -0.622    0.535
contextnot > most/many    3.832     10.406  16.172   0.368    0.717
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Correlation of Fixed Effects:
            (Intr) cnt/>n cntx>a
cntxtmst/>n -0.382
cntxtnt>all -0.338  0.399
cntxtn>mst/ -0.180  0.628  0.422
```

For the models above, there's no significant correlation between sentence context and fundamental frequecies in English.

Context vs Fundamental Frequency(F0), Japanese

back to top

```
model.minF0.Jap <- lmer(data = data.SenLen.Jap,
                        minF0 ~ context +
                          (1 + context||keyword) +
                          (1 + context||subject), REML = FALSE,
                        control=lmerControl(check.conv.singular = .makeCC(action = "ignore",  tol = 1e-
summary(model.minF0.Jap)
```

Model for the min fundamental frequency during the sentence (minF0):

```
Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
  method [lmerModLmerTest]
Formula: minF0 ~ context + (1 + context || keyword) + (1 + context ||
    subject)
   Data: data.SenLen.Jap
Control: lmerControl(check.conv.singular = .makeCC(action = "ignore",
    tol = 1e-04))

     AIC      BIC   logLik deviance df.resid
  2100.6   2189.6  -1023.3   2046.6      173

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.6233 -0.2898  0.2083  0.5220  1.5242

Random effects:
 Groups    Name                    Variance  Std.Dev.  Corr
 subject   contextall > not        1.592e+03 3.991e+01
           contextmost/many > not  1.334e+03 3.652e+01 0.97
           contextnot > all        1.693e+03 4.115e+01 1.00  0.94
           contextnot > most/many  2.124e+03 4.608e+01 0.99  0.99  0.98
 subject.1 (Intercept)             0.000e+00 0.000e+00
 keyword   contextall > not        1.914e+01 4.375e+00
           contextmost/many > not  2.343e-08 1.531e-04 -1.00
           contextnot > all        1.223e+00 1.106e+00  1.00 -1.00
           contextnot > most/many  6.865e-07 8.285e-04 -0.95  0.95 -0.95
 keyword.1 (Intercept)             0.000e+00 0.000e+00
 Residual                          1.219e+03 3.491e+01
```

```
Number of obs: 200, groups:  subject, 20; keyword, 5

Fixed effects:
                       Estimate Std. Error      df t value Pr(>|t|)
(Intercept)             134.536     10.311  15.515  13.048 8.96e-10 ***
contextmost/many > not   15.102      7.916  12.605   1.908   0.0794 .
contextnot > all          5.868      6.692  12.063   0.877   0.3977
contextnot > most/many   11.501      7.779  13.621   1.478   0.1621
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) cnt/>n cntx>a
cntxtmst/>n -0.440
cntxtnt>all -0.331  0.440
cntxtn>mst/ -0.188  0.456  0.470
```

```
model.maxF0.Jap <- lmer(data = data.SenLen.Jap,
                        maxF0 ~ context +
                          (1 + context||keyword) +
                          (1 + context||subject), REML = FALSE,
                        control=lmerControl(check.conv.singular = .makeCC(action = "ignore",  tol = 1e-
```

Model for the max fundamental frequency during the sentence (maxF0):

```
Warning: Model failed to converge with 1 negative eigenvalue: -8.8e-02
```

```
summary(model.maxF0.Jap)
```

```
Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
  method [lmerModLmerTest]
Formula: maxF0 ~ context + (1 + context || keyword) + (1 + context ||
    subject)
   Data: data.SenLen.Jap
Control: lmerControl(check.conv.singular = .makeCC(action = "ignore",
    tol = 1e-04))

     AIC      BIC   logLik deviance df.resid
  2100.2   2189.2  -1023.1   2046.2      173

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.4419 -0.4916 -0.0873  0.4028  3.7869

Random effects:
 Groups    Name                    Variance Std.Dev. Corr
 subject   contextall > not        1858.31  43.108
           contextmost/many > not  2291.80  47.873   0.99
           contextnot > all        2961.99  54.424   0.98  0.93
           contextnot > most/many  3035.66  55.097   0.99  0.96  0.99
```

```
 subject.1 (Intercept)               0.00   0.000
 keyword   contextall > not        136.85  11.698
           contextmost/many > not    5.05   2.247    1.00
           contextnot > all         32.99   5.744   -1.00 -1.00
           contextnot > most/many   29.70   5.450    1.00  1.00 -1.00
 keyword.1 (Intercept)               0.00   0.000
 Residual                         1114.19  33.379
Number of obs: 200, groups:  subject, 20; keyword, 5

Fixed effects:
                       Estimate Std. Error      df t value Pr(>|t|)
(Intercept)             304.742     12.534  12.588  24.313 5.86e-12 ***
contextmost/many > not   -7.520      9.914   6.572  -0.759    0.474
contextnot > all          6.095     12.261   3.520   0.497    0.649
contextnot > most/many   -6.628     10.815   5.734  -0.613    0.563
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) cnt/>n cntx>a
cntxtmst/>n -0.444
cntxtnt>all -0.422  0.694
cntxtn>mst/ -0.294  0.662  0.721
```

For the models above, there's no significant correlation between sentence context and fundamental frequecies in Japanese