Analysis code for Top-down versus bottom-up theories of phonological acquisition: A big data approach

Christina Bergmann, Sho Tsuji, Alejandrina Cristia

Introduction

This document provides the analysis code for the Interspeech 2017 paper "Top-down versus bottom-up theories of phonological acquisition: A big data approach".

Read-in and preprocessing steps can be found in the R Markdown document.

Analyses

All data

Vowel acquisition

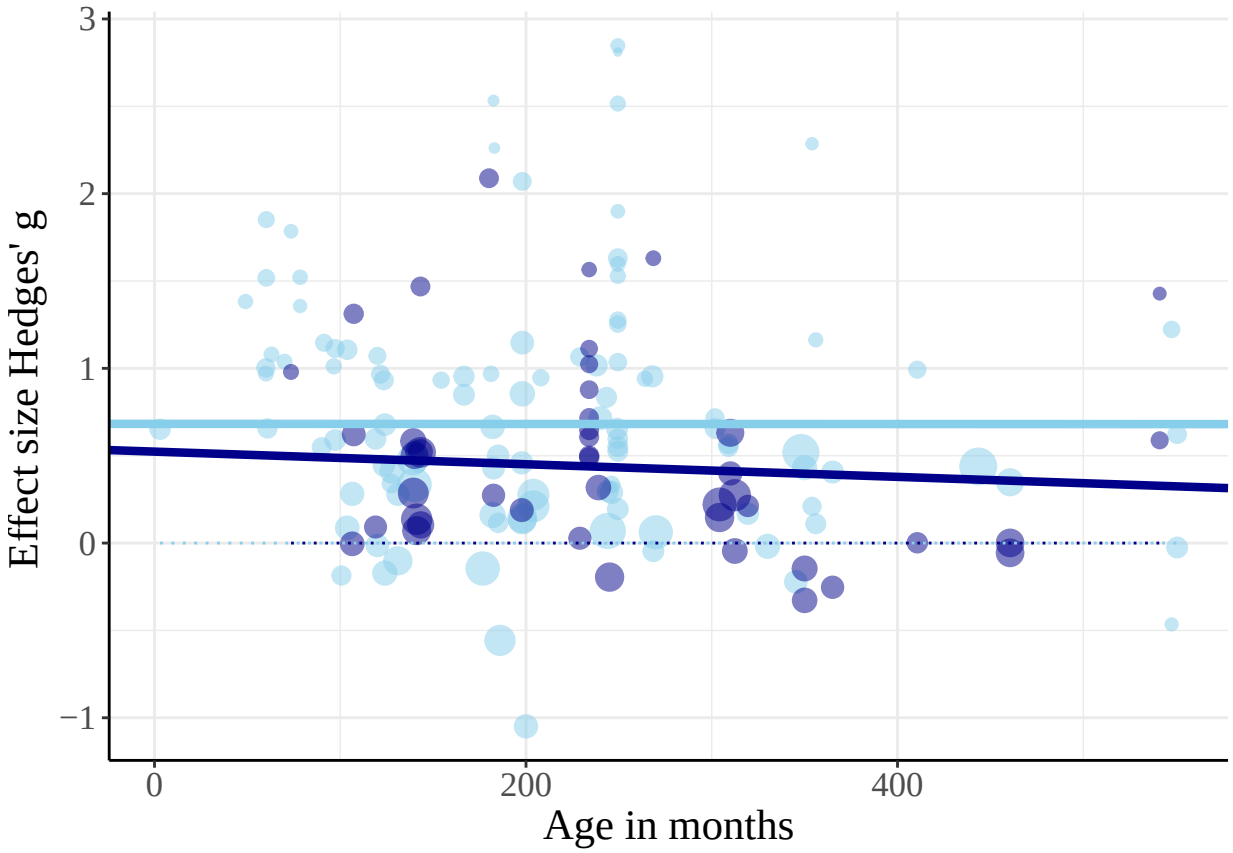
```
native.full<-rma.mv(g_calc, g_var_calc, mods=~ age.C + exposure_phase,random = ~ same_infant| short_cit
summary(native.full)
```

```
##
## Multivariate Meta-Analysis Model (k = 108; method: REML)
##
##   logLik  Deviance       AIC       BIC      AICc
## -89.3047  178.6095  192.6095  211.0526  193.7884
##
## Variance Components:
##
## outer factor: short_cite  (nlvls = 29)
## inner factor: same_infant (nlvls = 105)
##
##             estim    sqrt  fixed
## tau^2      0.2623  0.5121     no
## rho        0.4665             no
##
## Test for Residual Heterogeneity:
## QE(df = 103) = 345.4418, p-val < .0001
##
## Test of Moderators (coefficient(s) 2,3,4,5):
## QM(df = 4) = 4.3830, p-val = 0.3567
##
## Model Results:
##
##                  estimate      se     zval    pval    ci.lb   ci.ub
## intrcpt            0.6817  0.0983   6.9383  <.0001   0.4892  0.8743  ***
## age.C             -0.0001  0.0169  -0.0042  0.9966  -0.0331  0.0330
## exposure_phase1    0.2777  0.1602   1.7331  0.0831  -0.0363  0.5917    .
## exposure_phase2    0.0233  0.2148   0.1084  0.9137  -0.3977  0.4442
```

```
## exposure_phase3   -0.1695  0.1335  -1.2700  0.2041  -0.4311  0.0921
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
nonnative.full<-rma.mv(g_calc, g_var_calc, mods=~ age.C + exposure_phase,random = ~ same_infant | short_
summary(nonnative.full)
```

```
##
## Multivariate Meta-Analysis Model (k = 45; method: REML)
##
##   logLik  Deviance       AIC       BIC      AICc
## -13.7392   27.4785   41.4785   53.3006   44.9785
##
## Variance Components:
##
## outer factor: short_cite  (nlvls = 13)
## inner factor: same_infant (nlvls = 29)
##
##             estim    sqrt  fixed
## tau^2      0.1456  0.3816     no
## rho        0.8668             no
##
## Test for Residual Heterogeneity:
## QE(df = 40) = 67.0471, p-val = 0.0047
##
## Test of Moderators (coefficient(s) 2,3,4,5):
## QM(df = 4) = 13.7980, p-val = 0.0080
##
## Model Results:
##
##                  estimate      se     zval    pval    ci.lb   ci.ub
## intrcpt            0.4431  0.1384   3.2019  0.0014   0.1719  0.7143  **
## age.C             -0.0110  0.0177  -0.6210  0.5346  -0.0457  0.0237
## exposure_phase1    0.5588  0.1986   2.8136  0.0049   0.1695  0.9481  **
## exposure_phase2   -0.3160  0.3124  -1.0115  0.3118  -0.9282  0.2963
## exposure_phase3    0.1703  0.2322   0.7334  0.4633  -0.2848  0.6253
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



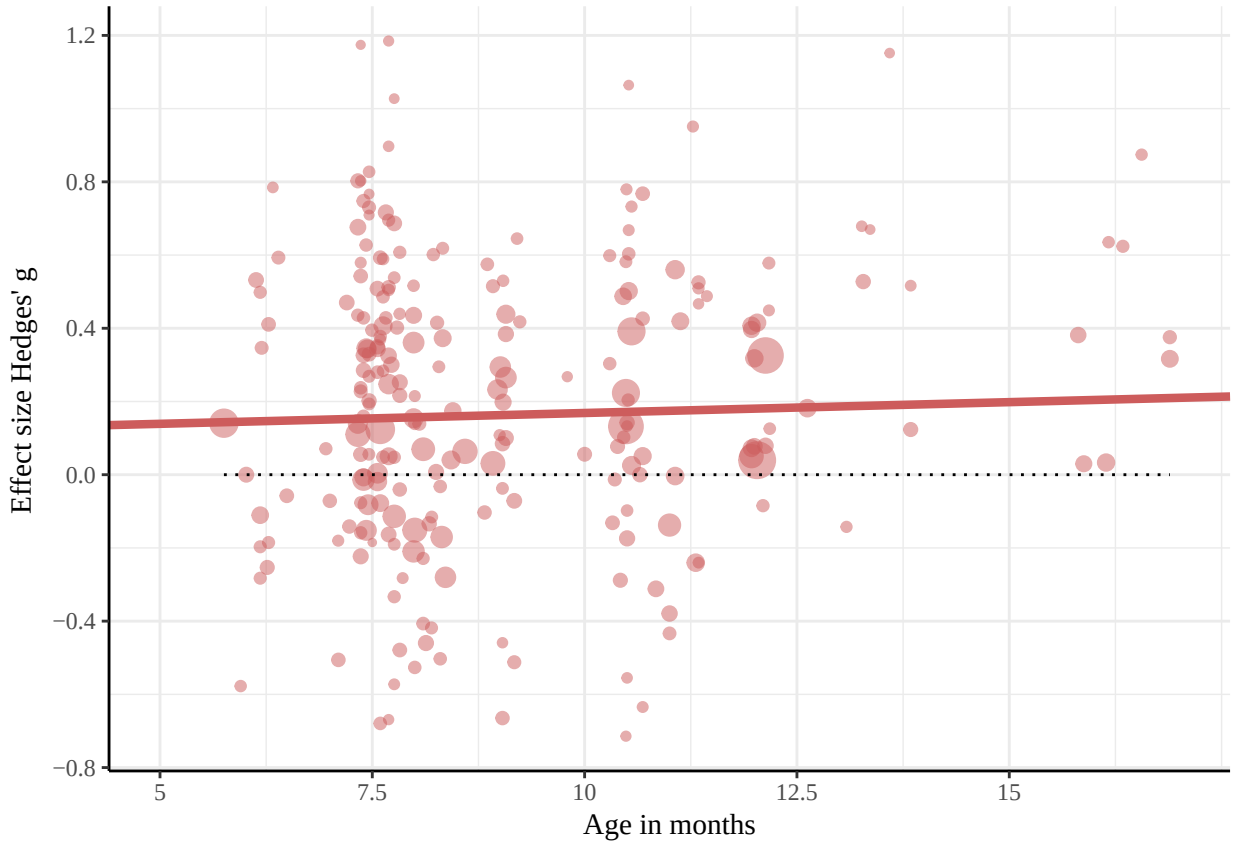
Word Segmentation

```
rma_age = rma.mv(g_calc, g_var_calc, mods = age.C,data = db, random =  ~ same_infant | short_cite)

summary(rma_age)
```

```
##
## Multivariate Meta-Analysis Model (k = 238; method: REML)
##
##    logLik   Deviance        AIC        BIC       AICc
## -128.2656   256.5312   264.5312   278.3865   264.7043
##
## Variance Components:
##
## outer factor: short_cite  (nlvls = 58)
## inner factor: same_infant (nlvls = 49)
##
##             estim    sqrt  fixed
## tau^2      0.0768  0.2771     no
## rho        0.1157             no
##
## Test for Residual Heterogeneity:
## QE(df = 236) = 871.9776, p-val < .0001
##
## Test of Moderators (coefficient(s) 2):
```

```
## QM(df = 1) = 0.5361, p-val = 0.4641
##
## Model Results:
##
##          estimate      se    zval    pval    ci.lb   ci.ub
## intrcpt    0.1647  0.0262  6.2787  <.0001   0.1133  0.2161  ***
## mods       0.0059  0.0081  0.7322  0.4641  -0.0099  0.0218
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



Only use data testing at least two age groups on same contrast

Vowels

Since nonnative model did not converge, dropping same_infant here! (Significant effect of age remains for both model with or without same_infant).

```
native.sub<-rma.mv(g_calc, g_var_calc, mods=~ age.C + exposure_phase,random = ~ same_infant | short_cit
summary(native.sub)
```

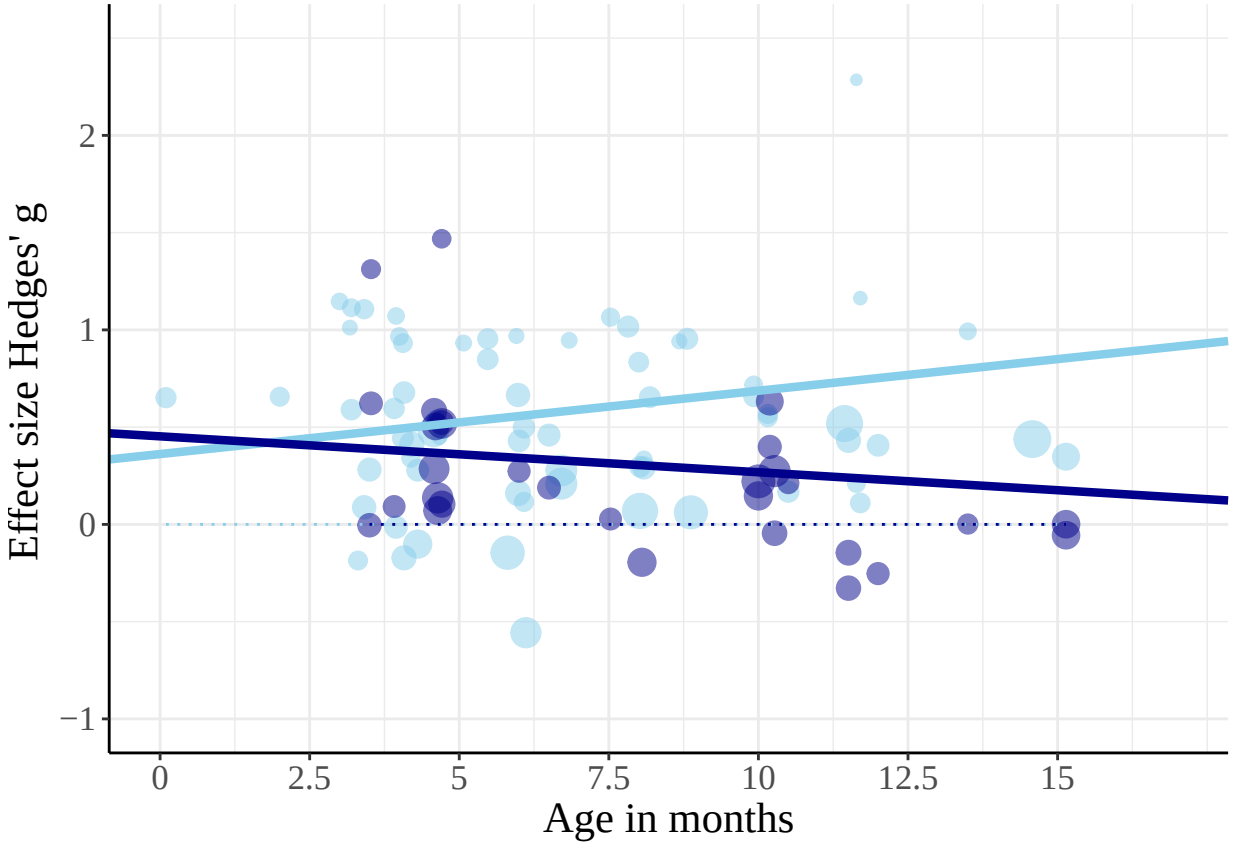
```
##
## Multivariate Meta-Analysis Model (k = 72; method: REML)
##
##   logLik  Deviance       AIC       BIC      AICc
## -39.3763   78.7525   92.7525  108.1854   94.6508
##
```

```
## Variance Components:
##
## outer factor: short_cite  (nlvls = 17)
## inner factor: same_infant (nlvls = 69)
##
##             estim    sqrt  fixed
## tau^2      0.1306  0.3614     no
## rho        0.6370             no
##
## Test for Residual Heterogeneity:
## QE(df = 67) = 164.1438, p-val < .0001
##
## Test of Moderators (coefficient(s) 2,3,4,5):
## QM(df = 4) = 13.6033, p-val = 0.0087
##
## Model Results:
##
##                  estimate      se     zval    pval    ci.lb    ci.ub
## intrcpt            0.5995  0.0974   6.1566  <.0001   0.4086   0.7903  ***
## age.C              0.0326  0.0148   2.1993  0.0279   0.0035   0.0616    *
## exposure_phase1    0.4022  0.2039   1.9727  0.0485   0.0026   0.8018    *
## exposure_phase2    0.1299  0.1814   0.7162  0.4738  -0.2256   0.4855
## exposure_phase3   -0.4567  0.1384  -3.2987  0.0010  -0.7280  -0.1853  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
nonnative.sub<-rma.mv(g_calc, g_var_calc, mods=~ age.C + exposure_phase,random = ~ 1 | short_cite, data=
summary(nonnative.sub)
```

```
##
## Multivariate Meta-Analysis Model (k = 30; method: REML)
##
##   logLik  Deviance       AIC       BIC      AICc
##  -1.0468    2.0936   14.0936   21.4068   18.7602
##
## Variance Components:
##
##             estim    sqrt  nlvls  fixed      factor
## sigma^2    0.0000  0.0000      7     no  short_cite
##
## Test for Residual Heterogeneity:
## QE(df = 25) = 30.4930, p-val = 0.2064
##
## Test of Moderators (coefficient(s) 2,3,4,5):
## QM(df = 4) = 35.4956, p-val < .0001
##
## Model Results:
##
##                  estimate      se     zval    pval    ci.lb   ci.ub
## intrcpt            0.3096  0.0544   5.6946  <.0001   0.2030  0.4161  ***
## age.C             -0.0158  0.0129  -1.2260  0.2202  -0.0411  0.0095
## exposure_phase1    0.5078  0.1187   4.2770  <.0001   0.2751  0.7405  ***
## exposure_phase2   -0.1946  0.1133  -1.7165  0.0861  -0.4167  0.0276    .
## exposure_phase3   -0.0365  0.0670  -0.5453  0.5856  -0.1679  0.0948
```

```
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



Word Segmentation

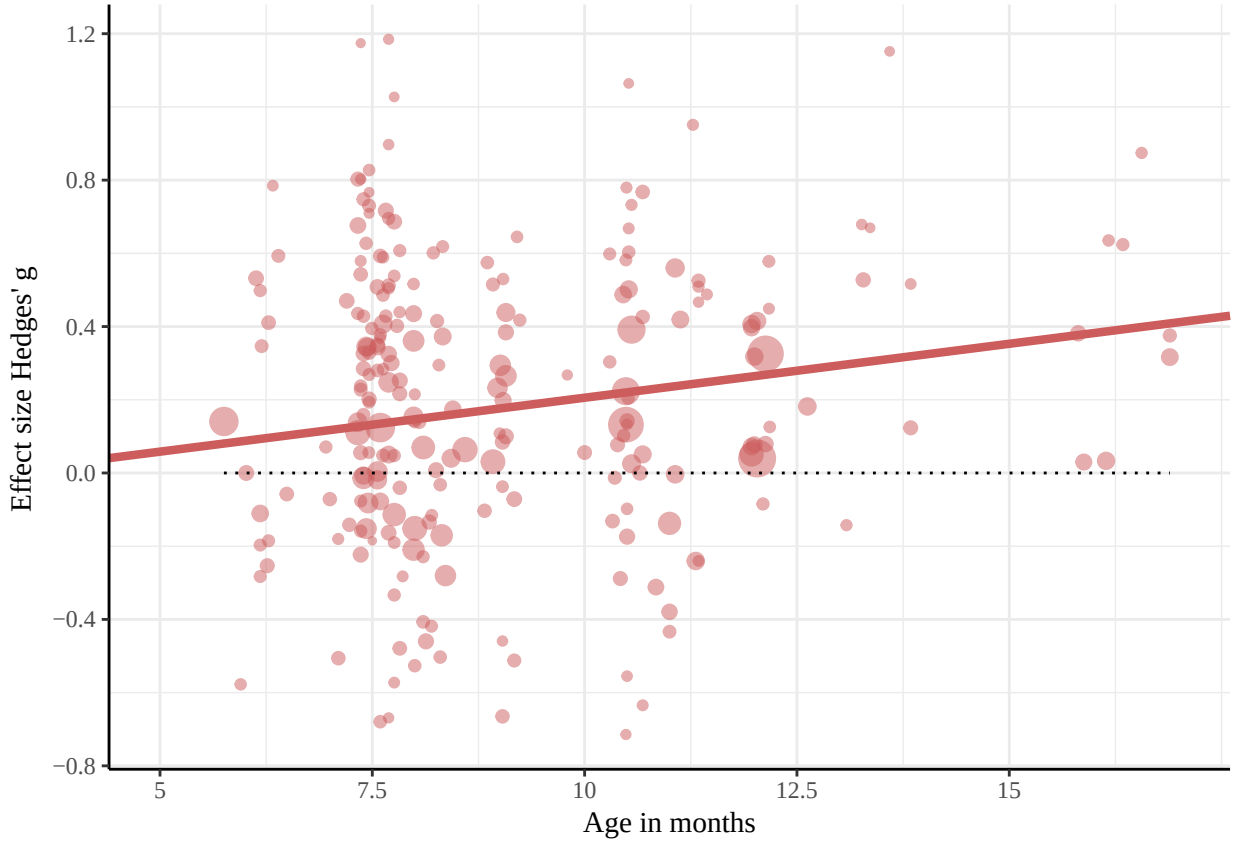
```
db_multiple = db %>%
  filter(multiple_age_groups == 1)

rma_age_m = rma.mv(g_calc, g_var_calc, data = db_multiple, mod = age.C, random =  ~ same_infant | short_

summary(rma_age_m)
```

```
##
## Multivariate Meta-Analysis Model (k = 80; method: REML)
##
##   logLik  Deviance       AIC       BIC      AICc
## -45.6674   91.3347   99.3347  108.7615   99.8826
##
## Variance Components:
##
## outer factor: short_cite  (nlvls = 22)
## inner factor: same_infant (nlvls = 30)
##
##             estim    sqrt  fixed
## tau^2      0.0952  0.3085     no
```

```
## rho        0.2927             no
##
## Test for Residual Heterogeneity:
## QE(df = 78) = 302.8673, p-val < .0001
##
## Test of Moderators (coefficient(s) 2):
## QM(df = 1) = 5.0539, p-val = 0.0246
##
## Model Results:
##
##          estimate      se    zval    pval   ci.lb   ci.ub
## intrcpt    0.1860  0.0552  3.3713  0.0007  0.0779  0.2942  ***
## mods       0.0295  0.0131  2.2481  0.0246  0.0038  0.0552    *
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



Joint figure