Supplementary Materials to: A thorough evaluation of the Language Environment Analysis (LENA) system

Alejandrina Cristia[1], Marvin Lavechin[1], Camila Scaff[1], Melanie Soderstrom[2], Caroline Rowland[3], Okko Räsänen[4,5], John Bunce[2,], & Elika Bergelson[6]

[1] Laboratoire de Sciences Cognitives et de Psycholinguistique, Département d'études cognitives, ENS, EHESS, CNRS, PSL University

[2] Department of Psychology, University of Manitoba, Canada

[3] Max Planck Institute for Psycholinguistics, Netherlands

[4] Unit of Computing Sciences, Tampere University, Finland

[5] Department of Signal Processing and Acoustics, Aalto University, Finland

[6] Psychology & Neuroscience, Duke University, Durham, North Carolina, USA

Author Note

Correspondence concerning this article should be addressed to Alejandrina Cristia, 29, rue d'Ulm, 75005 Paris, France. E-mail: alecristia@gmail.com

Abstract

This document provides the full results of analyses carried out in the main manuscript as well as other useful analyses.

Supplementary Materials to: A thorough evaluation of the Language Environment Analysis (LENA) system

LENA® classification accuracy: False alarms, misses, confusion

No additional results.

LENA® classification accuracy: Precision and recall

No additional results.

LENA® classification accuracy: Precision. No additional results.

LENA® classification accuracy: Recall. No additional results.

LENA® classification accuracy: Agreement using Cohen's kappa.

No additional results.

Derived metrics

Child Vocalization Counts (CVC) accuracy. In the present version of the main paper, we followed the LENA® definitions of what "counts" as a child vocalization for LENA®-reported CVC: Any vocalization that contains an utterance, i.e., a linguistic segment. Thus, child vocalizations consisting purely of vegetative/fix subsegments, or of cries, does not count towards CVC. In a previous version, however, we were counting any vocalization - regardless of its content. It turns out that accuracy is slightly better with the latter definition - which we will call CHN counts (so as to avoid a confusion with LENA®-reported Child Vocalization Counts).

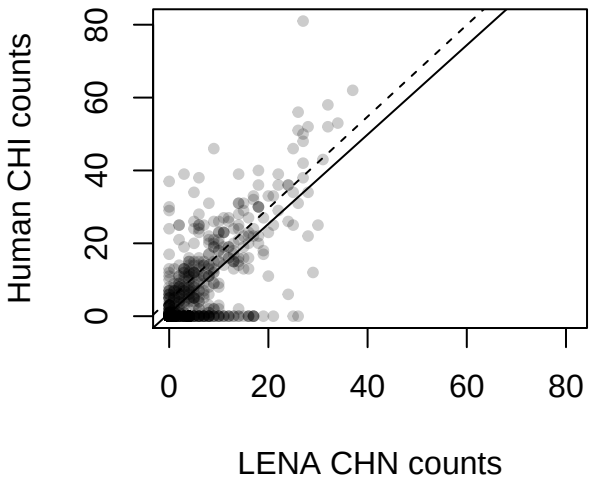

Figure 1. Key Child Segment counts recalculated from LENA reports (x axis; CHN counts, counting CHN segments regardless of whether they contain some linguistic stretches or not) and humans (y axis; CHI counts, counting segments attributed to the key child regardless of whether they were classified as linguistic or not). Each point represents the child segments totaled within a clip. The solid line corresponds to a linear regression fit to data from all clips; the dashed line corresponds to an analysis excluding clips where both the human and LENA® found zero child segments. The x and y ranges have been adjusted to be equal regardless of the data distribution.

Conversational Turn Counts (CTC) accuracy. As with CVC, LENA® counts infant-adult turns only if a child vocalization contains some linguistic material. In the main text, this is the analysis we report on, but we thought it informative to include in the present supplementary materials a redefinition whereby turns are counted regardless of whether the CHN segment contains some linguistic material or not. For instance, an infant cry followed by a child phrase would be counted in the latter definition but not the former.

Adult Word Counts accuracy. One of the children in the corpus was in a French-speaking environment. The following Figure shows results for AWC excluding the 15 data points corresponding to this child.

Table 1

For each measure (CVC, CHN counts, CTC, " all CHI, AWC, " no French – see main text for details), N all stands for the total number of clips considered in the general correlation analysis and resulting Pearson r coefficient; N stands for the number of non-null clips (i.e., having some vocalizations, turns, and adult words respectively) as well as Pearson r in the analyses restricted to the non-null clips.

	N clips	r all	N	r
CVC	757	0.641	255	0.638
CHN c	757	0.641	255	0.638
CTC	757	0.567	206	0.351
" all CHI	757	0.501	196	0.292
AWC	598	0.751	307	0.687
" no French	583	0.751	305	0.686

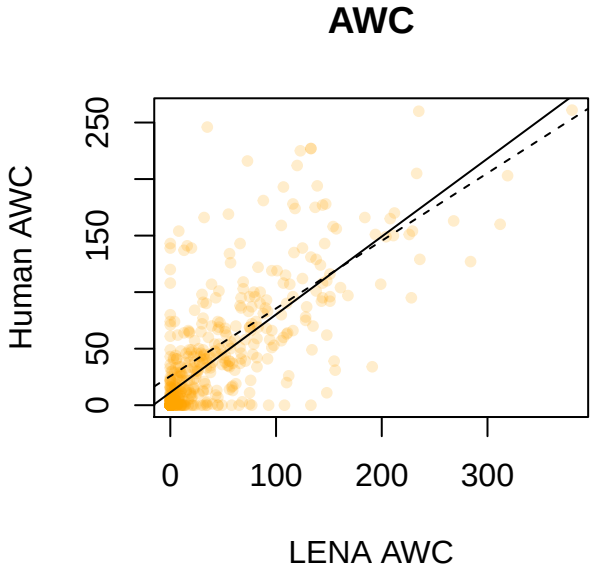

Table 2

For each measure (CVC, CHN counts, CTC, " all CHI, AWC, " no French – see main text for details), AER is the absolute error rate, AER-0 the same excluding clips with counts of zero according to either the human or the system, RER is the relative error rate and ARER the absolute relative error rate (the latter two exclude clips where the human count is zero).

	AER	min	max	AER-0	min	max	RER	min1	max1	ARER	min2	max2
CVC	-2	-37	29	-8	-35	14	-48	-100	650	76	0	650
CHN c	-2	-54	26	-7	-54	18	-28	-100	700	69	0	700
CTC	-2	-41	15	-5	-41	15	-29	-100	1,200	94	0	1,200
" all CHI	-2	-48	18	-6	-44	15	-27	-100	1,200	96	0	1,200
AWC	0	-211	157	0	-211	157	55	-100	7,400	124	0	7,400
" no French	0	-211	157	1	-211	157	56	-100	7,400	124	0	7,400

Effects of age and differences across corpora

Here we report on full model results.

Identification error rate.

```
## [1] "The following model corresponds to false.alarm.."
## Linear mixed model fit by REML ['lmerMod']
## Formula: py[, dv] ~ cor * age + (1 | child)
##    Data: py
##
## REML criterion at convergence: 10693.3
##
## Scaled residuals:
```

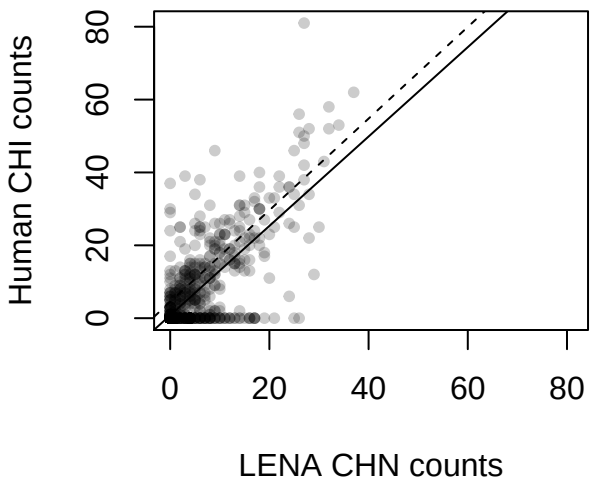

Figure 2. Conversational turns between Key Child and an adult recalculated from LENA reports and counting all CHN vocalizations, even if they do not contain any linguistic material (x axis) and humans (y axis; similarly counting all segments attributed to the key child regardless of whether they were classified as linguistic or not). Each point represents the turns totaled within a clip. The solid line corresponds to a linear regression fit to data from all clips; the dashed line corresponds to an analysis excluding clips where both the human and LENA® found zero turns. The x and y ranges have been adjusted to be equal regardless of the data distribution.

```
##     Min      1Q  Median      3Q     Max
## -1.9546 -0.1182 -0.0639 -0.0045 25.1886
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  child    (Intercept)  2542     50.42
##  Residual             23591    153.59
## Number of obs: 830, groups:  child, 53
##
## Fixed effects:
```

```
##             Estimate Std. Error t value
## (Intercept)  32.5213    70.9717   0.458
## corROW      -24.3956    99.3648  -0.246
## corSOD       50.9833    80.4927   0.633
## corTSI      -16.1728    84.3300  -0.192
## corWAR      -15.8109    89.8810  -0.176
## age          -1.4715     6.0724  -0.242
## corROW:age    1.8730     6.9236   0.271
## corSOD:age   -0.7689     6.6095  -0.116
## corTSI:age    1.8096     6.2003   0.292
## corWAR:age    1.1987    10.1558   0.118
##
## Correlation of Fixed Effects:
##             (Intr) corROW corSOD corTSI corWAR age    crROW: crSOD: crTSI:
## corROW      -0.714
## corSOD      -0.882  0.630
## corTSI      -0.842  0.601  0.742
## corWAR      -0.790  0.564  0.696  0.665
## age         -0.958  0.684  0.845  0.806  0.757
## corROW:age   0.840 -0.922 -0.741 -0.707 -0.664 -0.877
## corSOD:age   0.880 -0.629 -0.934 -0.741 -0.695 -0.919  0.806
## corTSI:age   0.939 -0.670 -0.827 -0.890 -0.741 -0.979  0.859  0.900
## corWAR:age   0.573 -0.409 -0.505 -0.482 -0.910 -0.598  0.524  0.549  0.586
## Analysis of Deviance Table (Type III Wald chisquare tests)
##
## Response: py[, dv]
##              Chisq Df Pr(>Chisq)
```

```
## (Intercept) 0.2100  1     0.6468
## cor         1.9768  4     0.7400
## age         0.0587  1     0.8085
## cor:age     0.8680  4     0.9291
## [1] "The following model corresponds to missed.detection.."
## Linear mixed model fit by REML ['lmerMod']
## Formula: py[, dv] ~ cor * age + (1 | child)
##    Data: py
##
## REML criterion at convergence: 8054.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.80065 -0.89369 -0.05148  0.73440  2.39729
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  child    (Intercept) 125.5    11.20
##  Residual             937.7    30.62
## Number of obs: 830, groups:  child, 53
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  37.79775   15.17127   2.491
## corROW       15.75985   21.24073   0.742
## corSOD        6.05640   17.20654   0.352
## corTSI      -11.93678   18.03492  -0.662
```

```
## corWAR        2.93300   19.21343   0.153
## age           0.18724    1.29807   0.144
## corROW:age   -1.08155    1.48002  -0.731
## corSOD:age   -0.08728    1.41289  -0.062
## corTSI:age    0.07516    1.32550   0.057
## corWAR:age    0.12401    2.17095   0.057
##
## Correlation of Fixed Effects:
##            (Intr) corROW corSOD corTSI corWAR age    crROW: crSOD: crTSI:
## corROW     -0.714
## corSOD     -0.882  0.630
## corTSI     -0.841  0.601  0.742
## corWAR     -0.790  0.564  0.696  0.664
## age        -0.958  0.684  0.845  0.806  0.757
## corROW:age  0.840 -0.922 -0.741 -0.707 -0.664 -0.877
## corSOD:age  0.880 -0.629 -0.934 -0.741 -0.695 -0.919  0.806
## corTSI:age  0.938 -0.670 -0.827 -0.890 -0.741 -0.979  0.859  0.900
## corWAR:age  0.573 -0.409 -0.505 -0.482 -0.910 -0.598  0.524  0.549  0.586
## Analysis of Deviance Table (Type III Wald chisquare tests)
##
## Response: py[, dv]
##              Chisq Df Pr(>Chisq)
## (Intercept) 6.2071  1    0.01272 *
## cor         3.1580  4    0.53174
## age         0.0208  1    0.88531
## cor:age     2.3331  4    0.67476
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] "The following model corresponds to confusion.."
## Linear mixed model fit by REML ['lmerMod']
## Formula: py[, dv] ~ cor * age + (1 | child)
##    Data: py
##
## REML criterion at convergence: 4677.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.9826 -0.5780 -0.2471  0.3239  5.5138
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  child    (Intercept)   0.0     0.00
##  Residual             197.1    14.04
## Number of obs: 578, groups:  child, 53
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept) 11.30672    5.11879   2.209
## corROW       1.80466    7.13277   0.253
## corSOD      -6.19432    5.61288  -1.104
## corTSI      -1.13895    6.00533  -0.190
## corWAR      -1.89357    6.20817  -0.305
## age         -0.15678    0.43563  -0.360
## corROW:age   0.08925    0.49994   0.179
```

```
## corSOD:age   0.86681    0.46069   1.882
## corTSI:age   0.29533    0.44353   0.666
## corWAR:age  -0.03466    0.67419  -0.051
##
## Correlation of Fixed Effects:
##            (Intr) corROW corSOD corTSI corWAR age    crROW: crSOD: crTSI:
## corROW     -0.718
## corSOD     -0.912  0.654
## corTSI     -0.852  0.612  0.777
## corWAR     -0.825  0.592  0.752  0.703
## age        -0.963  0.691  0.878  0.821  0.794
## corROW:age  0.839 -0.931 -0.765 -0.715 -0.692 -0.871
## corSOD:age  0.911 -0.654 -0.939 -0.776 -0.751 -0.946  0.824
## corTSI:age  0.946 -0.679 -0.863 -0.898 -0.780 -0.982  0.856  0.929
## corWAR:age  0.622 -0.447 -0.568 -0.530 -0.913 -0.646  0.563  0.611  0.635
## convergence code: 0
## boundary (singular) fit: see ?isSingular
##
## Analysis of Deviance Table (Type III Wald chisquare tests)
##
## Response: py[, dv]
##               Chisq Df Pr(>Chisq)
## (Intercept)  4.8791  1   0.027184 *
## cor          3.6533  4   0.454953
## age          0.1295  1   0.718932
## cor:age     14.3943  4   0.006137 **
## ---
```

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

There is a 2-way interaction between age and corpus for confusion. To investigate this we fit the same regression within each corpus.

```
## [1] "BER"
## Linear mixed model fit by REML ['lmerMod']
## Formula: py[, "confusion.."] ~ age + (1 | child)
##    Data: py
##  Subset: c(cor == thiscor)
## REML criterion at convergence: 778.9731
## Random effects:
##  Groups   Name        Std.Dev.
##  child    (Intercept)  0.00
##  Residual             10.41
## Number of obs: 104, groups:  child, 10
## Fixed Effects:
## (Intercept)          age
##     11.3067      -0.1568
## convergence code 0; 1 optimizer warnings; 0 lme4 warnings
## Analysis of Deviance Table (Type III Wald chisquare tests)
##
## Response: py[, "confusion.."]
##              Chisq Df Pr(>Chisq)
## (Intercept) 8.8740  1   0.002893 **
## age         0.2356  1   0.627427
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## [1] "ROW"
## Linear mixed model fit by REML ['lmerMod']
## Formula: py[, "confusion.."] ~ age + (1 | child)
##    Data: py
##  Subset: c(cor == thiscor)
## REML criterion at convergence: 866.1401
## Random effects:
##  Groups   Name        Std.Dev.
##  child    (Intercept)  1.27
##  Residual             15.19
## Number of obs: 105, groups:  child, 10
## Fixed Effects:
## (Intercept)          age
##    13.11500     -0.06799
## Analysis of Deviance Table (Type III Wald chisquare tests)
##
## Response: py[, "confusion.."]
##              Chisq Df Pr(>Chisq)
## (Intercept) 5.5778  1    0.01819 *
## age         0.0617  1    0.80383
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] "SOD"
## Linear mixed model fit by REML ['lmerMod']
## Formula: py[, "confusion.."] ~ age + (1 | child)
##    Data: py
##  Subset: c(cor == thiscor)
```

```
## REML criterion at convergence: 946.4319
## Random effects:
##  Groups   Name        Std.Dev.
##  child    (Intercept)  0.00
##  Residual             17.45
## Number of obs: 111, groups:  child, 10
## Fixed Effects:
## (Intercept)          age
##       5.112        0.710
## convergence code 0; 1 optimizer warnings; 0 lme4 warnings
## Analysis of Deviance Table (Type III Wald chisquare tests)
##
## Response: py[, "confusion.."]
##              Chisq Df Pr(>Chisq)
## (Intercept)  3.191  1  0.0740434 .
## age         14.532  1  0.0001378 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] "TSI"
## Linear mixed model fit by REML ['lmerMod']
## Formula: py[, "confusion.."] ~ age + (1 | child)
##    Data: py
##  Subset: c(cor == thiscor)
## REML criterion at convergence: 1203.745
## Random effects:
##  Groups   Name        Std.Dev.
##  child    (Intercept)  0
```

```
##  Residual             15
## Number of obs: 146, groups:  child, 13
## Fixed Effects:
## (Intercept)          age
##     10.1678       0.1385
## convergence code 0; 1 optimizer warnings; 0 lme4 warnings
## Analysis of Deviance Table (Type III Wald chisquare tests)
##
## Response: py[, "confusion.."]
##              Chisq Df Pr(>Chisq)
## (Intercept) 9.1868  1   0.002438 **
## age         2.4218  1   0.119654
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] "WAR"
## Linear mixed model fit by REML ['lmerMod']
## Formula: py[, "confusion.."] ~ age + (1 | child)
##    Data: py
##  Subset: c(cor == thiscor)
## REML criterion at convergence: 832.9831
## Random effects:
##  Groups   Name        Std.Dev.
##  child    (Intercept) 2.417
##  Residual             9.953
## Number of obs: 112, groups:  child, 10
## Fixed Effects:
## (Intercept)          age
```

```
##      9.5289      -0.1846
## Analysis of Deviance Table (Type III Wald chisquare tests)
##
## Response: py[, "confusion.."]
##              Chisq Df Pr(>Chisq)
## (Intercept) 8.5845  1    0.00339 **
## age         0.1494  1    0.69910
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

CVC.

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: gold_CV_count ~ lena_CV_count * age * cor + (1 | child)
##    Data: cvtc
##
## REML criterion at convergence: 4838.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.7042 -0.5065  0.0000  0.0779  5.3689
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  child    (Intercept)  2.403   1.550
##  Residual             31.449   5.608
## Number of obs: 759, groups:  child, 52
##
## Fixed effects:
##                           Estimate Std. Error t value
## (Intercept)               6.005141   2.553123   2.352
## lena_CV_count             1.234632   0.406472   3.037
## age                      -0.198742   0.219155  -0.907
## corROW                   -0.181301   3.664423  -0.049
## corSOD                   -5.513557   3.228029  -1.708
## corTSI                   -6.005141   3.019312  -1.989
## corWAR                    0.634258   3.247934   0.195
## lena_CV_count:age         0.001936   0.032899   0.059
```

```
## lena_CV_count:corROW     -0.473526   0.714095  -0.663
## lena_CV_count:corSOD      0.404189   0.467873   0.864
## lena_CV_count:corTSI     -1.234632   0.444172  -2.780
## lena_CV_count:corWAR      0.120646   0.447716   0.269
## age:corROW               -0.030194   0.253402  -0.119
## age:corSOD                0.426819   0.282938   1.509
## age:corTSI                0.198742   0.223738   0.888
## age:corWAR               -0.114576   0.369271  -0.310
## lena_CV_count:age:corROW  0.076737   0.052404   1.464
## lena_CV_count:age:corSOD -0.034899   0.039174  -0.891
## lena_CV_count:age:corTSI -0.001936   0.033295  -0.058
## lena_CV_count:age:corWAR -0.019487   0.042851  -0.455


## Analysis of Deviance Table (Type III Wald chisquare tests)
##
## Response: gold_CV_count
##                          Chisq Df Pr(>Chisq)
## (Intercept)             5.5323  1   0.018669 *
## lena_CV_count           9.2260  1   0.002386 **
## age                     0.8224  1   0.364484
## cor                    10.9022  4   0.027685 *
## lena_CV_count:age       0.0035  1   0.953072
## lena_CV_count:cor      42.0418  4  1.635e-08 ***
## age:cor                 6.4464  4   0.168198
## lena_CV_count:age:cor   6.4835  4   0.165833
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

There is a 3-way interaction between age, corpus, and the predictive value of the LENA® system's counts with respect to the gold counts. To investigate this we fit the same regression within each corpus.

CTC.

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: gold_CTC_count ~ lena_CTC_count * age * cor + (1 | child)
##    Data: cvtc
##
## REML criterion at convergence: 4696.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.4568 -0.4344 -0.0996  0.1008  6.9909
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  child    (Intercept)  3.021   1.738
##  Residual             25.881   5.087
## Number of obs: 759, groups:  child, 52
##
## Fixed effects:
##                           Estimate Std. Error t value
## (Intercept)                0.52017    2.52450   0.206
## lena_CTC_count             2.05115    0.46932   4.370
## age                        0.16014    0.21728   0.737
## corROW                     4.68582    3.61093   1.298
## corSOD                     2.27593    3.24347   0.702
## corTSI                    -0.33784    3.00803  -0.112
## corWAR                     0.22087    3.20439   0.069
## lena_CTC_count:age        -0.06229    0.04129  -1.509
```

```
## lena_CTC_count:corROW     -0.03654    0.77926  -0.047
## lena_CTC_count:corSOD     -1.30924    0.58063  -2.255
## lena_CTC_count:corTSI     -1.51425    0.54685  -2.769
## lena_CTC_count:corWAR     -0.77413    0.53003  -1.461
## age:corROW                -0.36594    0.25086  -1.459
## age:corSOD                -0.06901    0.28454  -0.243
## age:corTSI                -0.15482    0.22188  -0.698
## age:corWAR                 0.06088    0.36425   0.167
## lena_CTC_count:age:corROW  0.02065    0.06186   0.334
## lena_CTC_count:age:corSOD  0.09019    0.05214   1.730
## lena_CTC_count:age:corTSI  0.05870    0.04214   1.393
## lena_CTC_count:age:corWAR  0.01307    0.05860   0.223

## Analysis of Deviance Table (Type III Wald chisquare tests)
##
## Response: gold_CTC_count
##                          Chisq Df Pr(>Chisq)
## (Intercept)             0.0425  1    0.83675
## lena_CTC_count         19.1008  1   1.24e-05 ***
## age                     0.5432  1    0.46112
## cor                     3.4125  4    0.49131
## lena_CTC_count:age      2.2761  1    0.13138
## lena_CTC_count:cor     11.8988  4    0.01812 *
## age:cor                 4.0640  4    0.39741
## lena_CTC_count:age:cor  4.7499  4    0.31394
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

AWC.

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: gold ~ LENA * age * cor + (1 | child)
##    Data: awc
##
## REML criterion at convergence: 5879.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.5861 -0.4075 -0.1605  0.2618  5.9831
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  child    (Intercept) 196.1    14.00
##  Residual             962.6    31.03
## Number of obs: 600, groups:  child, 40
##
## Fixed effects:
##                 Estimate Std. Error t value
## (Intercept)      0.88050   18.79928   0.047
## LENA             1.15820    0.17198   6.734
## age              1.15032    1.60182   0.718
## corROW          28.62155   26.26054   1.090
## corSOD          11.09934   21.34535   0.520
## corWAR           8.91309   24.07148   0.370
## LENA:age        -0.04398    0.01370  -3.211
## LENA:corROW     -0.47377    0.26423  -1.793
```

```
## LENA:corSOD     -0.66574    0.19755  -3.370
## LENA:corWAR     -0.39316    0.22282  -1.765
## age:corROW      -2.01925    1.82319  -1.108
## age:corSOD      -1.42997    1.73943  -0.822
## age:corWAR      -1.25289    2.71354  -0.462
## LENA:age:corROW  0.05108    0.01706   2.993
## LENA:age:corSOD  0.06509    0.01773   3.671
## LENA:age:corWAR  0.02676    0.02443   1.096

## Analysis of Deviance Table (Type III Wald chisquare tests)
##
## Response: gold
##                Chisq Df Pr(>Chisq)
## (Intercept)   0.0022  1   0.962643
## LENA         45.3512  1  1.647e-11 ***
## age           0.5157  1   0.472675
## cor           1.2769  3   0.734634
## LENA:age     10.3113  1   0.001322 **
## LENA:cor     11.8133  3   0.008051 **
## age:cor       1.2533  3   0.740254
## LENA:age:cor 14.9656  3   0.001846 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

There is a 3-way interaction between age, corpus, and the predictive value of the LENA® system's counts with respect to the gold counts. To investigate this we fit the same regression within each corpus.