

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

Alejandrina Cristia¹, Marvin Lavechin¹, Camila Scaff¹, Melanie Soderstrom², Caroline
Rowland³, Okko Räsänen^{4,5}, John Bunce², & Erika Bergelson⁶

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).

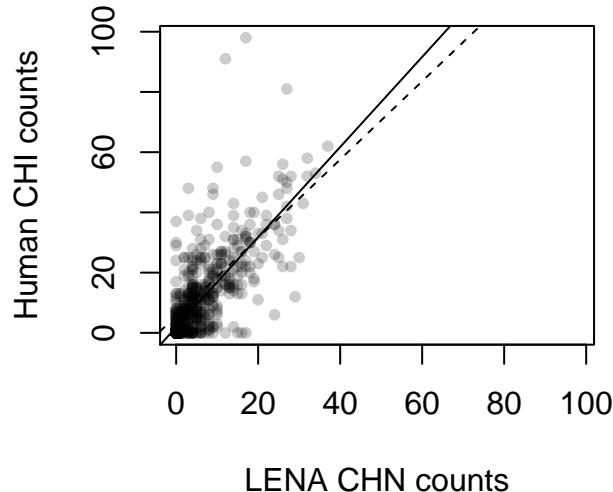


Figure 1. Key Child Segment counts recalculated from LENA reports (x axis; CHN counts, counting CHN segments regardless of whether they contain some linguistic stretches or not) and humans (y axis; CHI counts, counting segments attributed to the key child regardless of whether they were classified as 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	799	0.758	341	0.649
CHN c	799	0.758	341	0.649
CTC	799	0.572	201	0.369
" all CHI	799	0.513	196	0.296
AWC	598	0.751	307	0.687
" no French	583	0.751	305	0.686

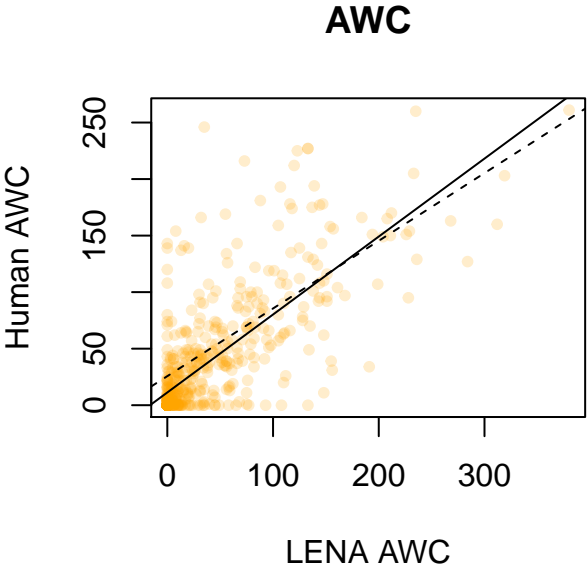


Table 2

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	-4	-38	14	-8	-38	14	-47	-100	650	76	0	650
CHN c	-4	-81	18	-8	-81	18	-32	-100	700	70	0	700
CTC	-2	-41	15	-5	-41	15	-35	-100	1,100	88	0	1,100
" all CHI	-2	-48	18	-6	-44	15	-32	-100	1,200	93	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: 10693.2
##
## Scaled residuals:
```

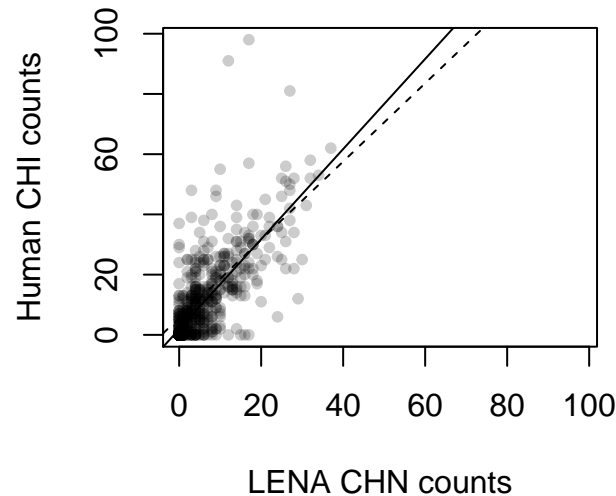


Figure 2. Conversational turns between Key Child and an adult recalculated from LENA reports and counting all CHN vocalizations, even if they do not contain any linguistic material (x axis) and humans (y axis; similarly counting all segments attributed to the key child regardless of whether they were 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
## -1.9550 -0.1182 -0.0639 -0.0045 25.1899
##
## Random effects:
## Groups   Name            Variance Std.Dev.
## child    (Intercept)    2543      50.43
## Residual                23588     153.58
## Number of obs: 830, groups: child, 53
##
## Fixed effects:
```

```

##              Estimate Std. Error t value
## (Intercept)  32.5236    70.9794   0.458
## corROW      -24.3919    99.3756  -0.245
## corSOD       50.9502    80.5015   0.633
## corTSI      -16.1744    84.3393  -0.192
## corWAR      -15.8131    89.8908  -0.176
## age         -1.4716     6.0731  -0.242
## corROW:age    1.8727     6.9243   0.270
## corSOD:age   -0.7699     6.6103  -0.116
## corTSI:age    1.8097     6.2010   0.292
## corWAR:age    1.1990    10.1569   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          1.9744  4      0.7405
## age          0.0587  1      0.8085
## cor:age      0.8685  4      0.9290
## [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
## -1.80062 -0.89360 -0.04501  0.73447  2.39722
##
## Random effects:
##  Groups   Name                Variance Std.Dev.
##  child    (Intercept) 125.4      11.20
##  Residual                937.7      30.62
## Number of obs: 830, groups:  child, 53
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  37.78949   15.16579   2.492
## corROW       15.76653   21.23305   0.743
## corSOD        6.06422   17.20032   0.353
## corTSI      -11.93022   18.02836  -0.662

```

```

## corWAR          2.93332   19.20648   0.153
## age             0.18806    1.29760   0.145
## corROW:age     -1.08224    1.47949  -0.731
## corSOD:age     -0.08763    1.41238  -0.062
## corTSI:age      0.07437    1.32502   0.056
## corWAR:age      0.12418    2.17017   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) 6.2089 1 0.01271 *
## cor          3.1605 4 0.53134
## age          0.0210 1 0.88477
## cor:age      2.3345 4 0.67450
## ---

```

```

## 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: 4677.6

##

## Scaled residuals:

##      Min       1Q   Median       3Q      Max
## -1.9822 -0.5781 -0.2476  0.3239  5.5146

##

## Random effects:

## Groups   Name      Variance Std.Dev.
## child    (Intercept)    0      0.00
## Residual                197     14.04

## Number of obs: 578, groups:  child, 53

##

## Fixed effects:

##              Estimate Std. Error t value
## (Intercept) 11.30495    5.11800   2.209
## corROW       1.80397    7.13167   0.253
## corSOD       -6.18712    5.61202  -1.102
## corTSI       -1.13718    6.00440  -0.189
## corWAR       -1.88840    6.20721  -0.304
## age          -0.15663    0.43557  -0.360
## corROW:age    0.08922    0.49986   0.178

```

```

## corSOD:age    0.86617    0.46062    1.880
## corTSI:age    0.29518    0.44347    0.666
## corWAR:age   -0.03531    0.67408   -0.052
##
## 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          3.6472  4  0.455849
## age          0.1293  1  0.719144
## cor:age      14.3776  4  0.006183 **
## ---

```

```
## 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: 5276.3
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.3559 -0.5212 -0.1493  0.2710  4.8466
##
## Random effects:
## Groups Name Variance Std.Dev.
## child (Intercept) 2.317 1.522
## Residual 39.464 6.282
## Number of obs: 801, groups: child, 52
##
## Fixed effects:
##
```

	Estimate	Std. Error	t value
(Intercept)	5.992098	2.704772	2.215
lena_CV_count	1.240664	0.453347	2.737
age	-0.198115	0.232310	-0.853
corROW	-0.174831	3.893447	-0.045
corSOD	-5.505476	3.423558	-1.608
corTSI	-5.196406	3.138631	-1.656
corWAR	0.599659	3.442060	0.174
lena_CV_count:age	0.001601	0.036733	0.044

```
## lena_CV_count:corROW      -0.458468    0.793713   -0.578
## lena_CV_count:corSOD       0.399488    0.522316    0.765
## lena_CV_count:corTSI       0.054959    0.493745    0.111
## lena_CV_count:corWAR       0.125204    0.499215    0.251
## age:corROW                 -0.030182    0.269053   -0.112
## age:corSOD                 0.426778    0.300182    1.422
## age:corTSI                 0.197628    0.236563    0.835
## age:corWAR                 -0.107948    0.391658   -0.276
## lena_CV_count:age:corROW   0.075733    0.058261    1.300
## lena_CV_count:age:corSOD  -0.034727    0.043772   -0.793
## lena_CV_count:age:corTSI   0.015773    0.037152    0.425
## lena_CV_count:age:corWAR  -0.020740    0.047816   -0.434
```

```
## Analysis of Deviance Table (Type III Wald chisquare tests)
```

```
##
```

```
## Response: gold_CV_count
```

```
##               Chisq Df Pr(>Chisq)
## (Intercept)      4.9079  1  0.026734 *
## lena_CV_count      7.4894  1  0.006206 **
## age               0.7273  1  0.393769
## cor               8.4198  4  0.077355 .
## lena_CV_count:age  0.0019  1  0.965239
## lena_CV_count:cor  2.1250  4  0.712788
## age:cor           5.6670  4  0.225439
## lena_CV_count:age:cor 7.5304  4  0.110375
```

```
## ---
```

```
## 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: 4916.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.5388 -0.4363 -0.0692  0.0617  7.1571
##
## Random effects:
## Groups Name Variance Std.Dev.
## child (Intercept) 2.939 1.714
## Residual 24.649 4.965
## Number of obs: 801, groups: child, 52
##
## Fixed effects:
##
```

	Estimate	Std. Error	t value
(Intercept)	0.51914	2.47907	0.209
lena_CTC_count	2.05179	0.45811	4.479
age	0.16023	0.21335	0.751
corROW	4.68410	3.54507	1.321
corSOD	2.27757	3.18419	0.715
corTSI	-0.61732	2.90800	-0.212
corWAR	0.23974	3.14662	0.076
lena_CTC_count:age	-0.06235	0.04030	-1.547

```

## lena_CTC_count:corROW      -0.03855      0.76091  -0.051
## lena_CTC_count:corSOD      -1.31010      0.56675  -2.312
## lena_CTC_count:corTSI      -1.40114      0.53120  -2.638
## lena_CTC_count:corWAR      -0.77478      0.51736  -1.498
## age:corROW                  -0.36595      0.24629  -1.486
## age:corSOD                  -0.06909      0.27934  -0.247
## age:corTSI                  -0.15094      0.21749  -0.694
## age:corWAR                   0.06041      0.35766   0.169
## lena_CTC_count:age:corROW   0.02082      0.06040   0.345
## lena_CTC_count:age:corSOD   0.09023      0.05090   1.773
## lena_CTC_count:age:corTSI   0.05540      0.04111   1.348
## lena_CTC_count:age:corWAR   0.01296      0.05720   0.227

## Analysis of Deviance Table (Type III Wald chisquare tests)
##
## Response: gold_CTC_count
##
##              Chisq Df Pr(>Chisq)
## (Intercept)    0.0439  1    0.83413
## lena_CTC_count 20.0596  1 7.506e-06 ***
## age            0.5640  1    0.45264
## cor            3.9477  4    0.41313
## lena_CTC_count:age 2.3933  1    0.12186
## lena_CTC_count:cor 11.0553  4    0.02595 *
## age:cor         4.2535  4    0.37278
## lena_CTC_count:age:cor 4.7366  4    0.31541
## ---
## 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.