Peer speech project

Out of the lab corpora: Linaza, Vila, SerraSole, Marrero, Aguirre, OreaPine, Nieva, Ornat, Romero, Geneva, Pauline, Champaud, York, Leveille, Clark, Brown, Kuczaj, Providence, Sachs, Hall, Snow, Valian, Post, Gleason, Suppes, Braunwald, Bloom70, Caroline, Wagner, Rigol, Miller, Leo, Gaeltacht, ZhouDinner, TCCM, Beijing, LeeWongLeung, Utrecht, Wijnen, Stellenbosch, Thomas, Forrester, Wells, MPI-EVA-Manchester, Lara, Belfast, SCECL, Kovacevic, Tanja, Protassova, Antelmi, Calambrone, Klammler, Roma, Florianopolis, Santos, Jordina, Julia, MireiaEvaPascual, Avram, Ishii, Miyata, MiiPro, Hamasaki, Plunkett, Kari, Ringstad, Jiwon, Ryu, Jakarta, Demuth, Soto, Bodor, Reger, Vija, Beek, Korgesaar, Zupping, Argus, Kohler, Narasimhan, Doukas, Levy, BatEl, Ravid, BSF, Samadi, Family

Number of out of the lab corpora: 88

Role tags in all CHILDES : Target_Child, Father, Brother, Mother, Investigator, Adult, Unidentified, Observer, Sister, Child, Girl, Aunt, Playmate, Grandmother, Uncle, Family_Friend, Grandfather, Visitor, Cousin, Boy, Camera_Operator, Babysitter, Teenager, Toy, Environment, Non_Human, Student, Teacher, Sibling, Housekeeper, Media, Doctor, Group, Caretaker, Speaker, Nurse, Target_Adult

```
#finds corpora with peer speech
namePeerSpeech=list()
dataPeer=list()

#3 CHECK: all cousins, boys etc Nnot adults?
childSpeakers=c("Sister", "Brother", "Playmate", "Teenager", "Cousin", "Child", "Girl",
"Sibling", "Boy")

cSelectPeer=subset(cSelect, (cSelect$role %in% childSpeakers))
peerCorpusName=unique(cSelectPeer$corpus_name)
```

Out of lab CHILDES corpora with child speech (tags= Sister, Brother, Playmate, Teenager, Cousin, Child): Linaza, Vila, SerraSole, Marrero, Aguirre, Romero, Geneva, Pauline, Champaud, York, Clark, Brown, Kuczaj, Providence, Sachs, Hall, Valian, Post, Gleason, Suppes, Braunwald, Bloom70, Caroline, Wagner, Rigol, Miller, Leo, Gaeltacht, ZhouDinner, TCCM, Beijing, LeeWongLeung, Stellenbosch, Forrester, Wells, MPI-EVA-Manchester, Lara, Belfast, SCECL, Kovacevic, Calambrone, Santos, Jordina, MireiaEvaPascual, Ishii, Miyata, MiiPro, Hamasaki, Plunkett, Kari, Ryu, Jakarta, Demuth, Soto, Bodor, Reger, Vija, Korgesaar, Zupping, Argus, Kohler, Levy, BatEl, Ravid, BSF, Samadi, Family

Number of CHILDES corpora with peer speech: 67

```
# counts number of utterances per speaker for the selected out-of-lab corpora
#This chunk takes time to compute!
tablep=data.frame()

i=1
for (name in peerCorpusName[1:5]) {   #choosing only first 5 corpora to have it compile
 faster!
  cp<-get_utterances(corpus=name)
  tabletmp_<-cp %>% group_by(speaker_role) %>% summarise(no_rows = length(speaker_role))
  tablep<-rbind(tablep, tabletmp_)
  i=i+1}
tablep
```

```
## # A tibble: 47 x 2
##    speaker_role no_rows
##    <chr>          <int>
##  1 Adult            127
##  2 Brother          338
##  3 Father          2616
##  4 Investigator      10
##  5 Mother           382
##  6 Target_Child    2631
##  7 Unidentified      31
##  8 Adult           1619
##  9 Child            651
## 10 Father          2161
## # … with 37 more rows
```

```
nuttsSummary<-tablep %>% group_by(speaker_role) %>% summarise(no_rows = sum(no_rows))
nuttsSummary
```

```
## # A tibble: 18 x 2
##    speaker_role  no_rows
##    <chr>            <int>
##  1 Adult             2228
##  2 Aunt                82
##  3 Brother           2978
##  4 Child             5213
##  5 Family_Friend      384
##  6 Father           12954
##  7 Girl                 3
##  8 Grandfather        685
##  9 Grandmother       1778
## 10 Investigator      7592
## 11 Mother           36468
## 12 Nurse            10516
## 13 Observer         16936
## 14 Playmate             4
## 15 Sister            3145
## 16 Target_Child     67269
## 17 Uncle               22
## 18 Unidentified        31
```

Per corpus Number of utterances per speaker for CHILDES corpora with peer speech: (see table) Total Number of utterances per speaker for CHILDES corpora with peer speech: (see nuttsSummary)

```
# counts number of utterances per speaker for the selected out-of-lab corpora with wirel
ess recordings

tablew=data.frame()

#4 CHECK if only these
wirelessCorpusName<-c("Wells", "Demuth", "Hall")

i=1
for (name in wirelessCorpusName) {
  cw<-get_utterances(corpus=name)
  tabletmp<-cw %>% group_by(speaker_role) %>% summarise(no_rows = length(speaker_role))
  tablew<-rbind(tablew, tabletmp)
  i=i+1}
tablew
```

```
## # A tibble: 45 x 2
##    speaker_role  no_rows
##    <chr>           <int>
##  1 Adult            1082
##  2 Aunt              163
##  3 Brother           261
##  4 Child            8706
##  5 Cousin             10
##  6 Doctor              5
##  7 Family_Friend      95
##  8 Father           4033
##  9 Grandfather       314
## 10 Grandmother      1969
## # … with 35 more rows
```

```
nuttsWirelessSummary<-tablew %>% group_by(speaker_role) %>% summarise(no_rows = sum(no_r
ows))
nuttsWirelessSummary
```

```
## # A tibble: 22 x 2
##    speaker_role  no_rows
##    <chr>           <int>
##  1 Adult            7651
##  2 Aunt              163
##  3 Brother         17668
##  4 Child           41497
##  5 Cousin           8275
##  6 Doctor              5
##  7 Family_Friend      95
##  8 Father          14848
##  9 Grandfather       375
## 10 Grandmother     10063
## # … with 12 more rows
```

Per corpus Number of utterances per speaker for CHILDES corpora with peer speech AND wireless recordings: (see tablew) Total Number of utterances per speaker for CHILDES corpora with peer speech AND wireless recordings: (see nuttsWirelessSummary)

```
#reads demnuth corpus and counts utterances per speaker
demuth<-read.csv(file="/Users/lscpuser/Downloads/Copy_sesotho_CDI.csv", header=TRUE)
#demuth_noTargetChild<-subset(demuth, !(demuth$speaker_role_raw == "Target_Child" ))
sesotho_speakers<- demuth %>% group_by(speaker_role_raw) %>% summarise(no_rows = length
(speaker_role_raw))
sesotho_speakers<- sesotho_speakers %>%   arrange(desc(no_rows))
sesotho_speakers
```

```
## # A tibble: 13 x 2
##    speaker_role_raw no_rows
##    <fct>              <int>
##  1 Target_Child       27658
##  2 Mother              9638
##  3 Cousin              7561
##  4 Brother             5422
##  5 Investigator        5059
##  6 Grandmother         4950
##  7 Playmate            3345
##  8 Adult                915
##  9 Uncle                377
## 10 Sister               252
## 11 Father               180
## 12 Teenager             142
## 13 ""                    35
```

```
#reads %dir annotated demnuth corpus and counts utterances per speaker_id, matches with
 speaker role
sesotho_directed<- as.data.frame(demuth %>% group_by(childdirected) %>% summarise(no_row
s = length(childdirected)))
```

```
## Warning: Factor `childdirected` contains implicit NA, consider using
## `forcats::fct_explicit_na`
```

```
speakers_info<- as.data.frame(unique(demuth %>%select(speaker_id, speaker_role_raw, spea
ker_age, speaker_gender)))
sesotho_directed_si<-merge(x=sesotho_directed, y=speakers_info, by.x="childdirected", b
y.y="speaker_id", all.x)
sesotho_directed_si<- sesotho_directed_si %>%   arrange(desc(no_rows))
sesotho_directed_si
```

```
##     childdirected no_rows speaker_role_raw speaker_age speaker_gender
## 1            <NA>   48401                                           
## 2           16434      60           Cousin       4;5.0           <NA>
## 3           16502      37           Cousin       4;9.0           <NA>
## 4           16455      34      Grandmother        <NA>         Female
## 5           16711      32         Playmate        <NA>           <NA>
## 6           16631      31           Cousin       5;0.0           <NA>
## 7           16713      31           Cousin      17;0.0           <NA>
## 8           16747      28          Brother       4;5.0           Male
## 9           16588      27           Cousin      4;11.0           <NA>
## 10          16399      25      Grandmother        <NA>         Female
## 11          16460      25           Cousin       4;6.0           <NA>
## 12          16509      25           Cousin       4;9.0           <NA>
## 13          16494      24           Cousin       4;8.0           <NA>
## 14          16551      24           Cousin       5;3.0           <NA>
## 15          16575      24         Playmate        <NA>           <NA>
## 16          16730      24          Brother       8;0.0           Male
## 17          16501      23           Cousin       8;4.0           <NA>
## 18          16478      22           Cousin       8;3.0           <NA>
## 19          16525      22           Cousin       4;9.0           <NA>
## 20          16401      21     Investigator        <NA>           <NA>
## 21          16473      21           Cousin       4;8.0           <NA>
## 22          16486      20           Cousin       4;8.0           <NA>
## 23          16610      20           Cousin      4;11.0           <NA>
## 24          16422      19           Cousin       4;5.0           <NA>
## 25          16491      19         Teenager      15;0.0           <NA>
## 26          16569      19           Cousin      4;10.0           <NA>
## 27          16737      18           Mother        <NA>         Female
## 28          16437      17      Grandmother        <NA>         Female
## 29          16632      17     Investigator        <NA>           <NA>
## 30          16673      17           Cousin       5;1.0           <NA>
## 31          16405      16           Cousin       4;5.0           <NA>
## 32          16435      16           Cousin       8;0.0           <NA>
## 33          16710      16            Adult        <NA>           <NA>
## 34          16714      16           Mother        <NA>         Female
## 35          16731      16           Mother        <NA>         Female
## 36          16740      16          Brother       4;5.0           Male
## 37          16453      15           Cousin       4;6.0           <NA>
## 38          16620      15     Investigator        <NA>           <NA>
## 39          16677      15         Playmate        <NA>           <NA>
## 40          16429      14      Grandmother        <NA>         Female
## 41          16449      14      Grandmother        <NA>         Female
## 42          16472      14           Cousin       8;3.0           <NA>
## 43          16480      14           Cousin       4;8.0           <NA>
## 44          16584      14      Grandmother        <NA>         Female
## 45          16668      14           Cousin       5;1.0           <NA>
## 46          16748      14           Cousin      17;0.0           <NA>
## 47          16469      13     Investigator        <NA>           <NA>
## 48          16511      13           Cousin       8;4.0           <NA>
## 49          16617      13           Cousin      4;11.0           <NA>
## 50          16428      12            Adult        <NA>           <NA>
## 51          16451      12     Investigator        <NA>           <NA>
## 52          16465      12           Cousin       4;6.0           <NA>
```

```
## 53          16524      12          Cousin       8;4.0          <NA>
## 54          16540      12          Cousin       5;3.0          <NA>
## 55          16573      12          Mother        <NA>        Female
## 56          16702      12      Grandmother       <NA>        Female
## 57          16709      12     Investigator       <NA>          <NA>
## 58          16741      12          Cousin      17;0.0          <NA>
## 59          16745      12     Investigator       <NA>          <NA>
## 60          16414      11          Cousin       4;5.0          <NA>
## 61          16443      11          Cousin       4;5.0          <NA>
## 62          16481      11      Grandmother       <NA>        Female
## 63          16500      11          Mother        <NA>        Female
## 64          16574      11          Cousin      4;10.0          <NA>
## 65          16623      11          Cousin       5;0.0          <NA>
## 66          16661      11     Investigator       <NA>          <NA>
## 67          16676      11          Cousin       5;1.0          <NA>
## 68          16756      11          Mother        <NA>        Female
## 69          16415      10          Cousin       8;0.0          <NA>
## 70          16461      10      Grandmother       <NA>        Female
## 71          16520      10      Grandmother       <NA>        Female
## 72          16612      10      Grandmother       <NA>        Female
## 73          16431       9     Investigator       <NA>          <NA>
## 74          16448       9          Cousin       4;6.0          <NA>
## 75          16518       9          Cousin       4;9.0          <NA>
## 76          16666       9           Adult       <NA>          <NA>
## 77          16671       9           Adult       <NA>          <NA>
## 78          16724       9          Mother        <NA>        Female
## 79          16433       8          Mother        <NA>        Female
## 80          16444       8      Grandmother       <NA>        Female
## 81          16459       8          Cousin       8;1.0          <NA>
## 82          16470       8        Teenager      15;0.0          <NA>
## 83          16586       8     Investigator       <NA>          <NA>
## 84          16627       8         Playmate       <NA>          <NA>
## 85          16646       8          Cousin       5;0.0          <NA>
## 86          16663       8          Cousin       5;1.0          <NA>
## 87          16670       8     Investigator       <NA>          <NA>
## 88          16485       7          Cousin       8;3.0          <NA>
## 89          16532       7         Playmate       <NA>          <NA>
## 90          16583       7          Cousin      4;10.0          <NA>
## 91          16639       7           Adult       <NA>          <NA>
## 92          16650       7          Cousin       5;0.0          <NA>
## 93          16698       7          Cousin       5;4.0          <NA>
## 94          16722       7          Cousin      17;0.0          <NA>
## 95          16754       7         Brother       4;5.0          Male
## 96          16439       6     Investigator       <NA>          <NA>
## 97          16440       6          Mother        <NA>        Female
## 98          16487       6           Uncle       <NA>          Male
## 99          16495       6           Uncle       <NA>          Male
## 100         16543       6     Investigator       <NA>          <NA>
## 101         16552       6          Father        <NA>          Male
## 102         16559       6          Cousin      4;10.0          <NA>
## 103         16606       6     Investigator       <NA>          <NA>
## 104         16645       6          Mother        <NA>        Female
## 105         16649       6          Mother        <NA>        Female
```

```
## 106         16656       6     Investigator        <NA>          <NA>
## 107         16717       6     Investigator        <NA>          <NA>
## 108         16410       5     Investigator        <NA>          <NA>
## 109         16441       5           Cousin       8;0.0          <NA>
## 110         16488       5      Grandmother        <NA>        Female
## 111         16513       5      Grandmother        <NA>        Female
## 112         16562       5         Playmate        <NA>          <NA>
## 113         16593       5      Grandmother        <NA>        Female
## 114         16609       5            Adult        <NA>          <NA>
## 115         16629       5           Mother        <NA>        Female
## 116         16635       5     Investigator        <NA>          <NA>
## 117         16640       5           Cousin       5;0.0          <NA>
## 118         16642       5      Grandmother        <NA>        Female
## 119         16742       5           Sister      10;0.0        Female
## 120         16762       5          Brother       4;5.0          Male
## 121         16402       4           Mother        <NA>        Female
## 122         16406       4           Cousin       8;0.0          <NA>
## 123         16432       4         Playmate        <NA>          <NA>
## 124         16457       4     Investigator        <NA>          <NA>
## 125         16467       4      Grandmother        <NA>        Female
## 126         16471       4           Mother        <NA>        Female
## 127         16477       4           Mother        <NA>        Female
## 128         16546       4           Cousin       5;3.0          <NA>
## 129         16567       4           Mother        <NA>        Female
## 130         16600       4           Cousin      4;11.0          <NA>
## 131         16604       4      Grandmother        <NA>        Female
## 132         16608       4           Mother        <NA>        Female
## 133         16667       4            Adult        <NA>          <NA>
## 134         16675       4     Investigator        <NA>          <NA>
## 135         16678       4      Grandmother        <NA>        Female
## 136         16691       4           Cousin       5;4.0          <NA>
## 137         16725       4           Mother        <NA>        Female
## 138         16744       4           Mother        <NA>        Female
## 139         16395       3     Investigator        <NA>          <NA>
## 140         16403       3         Playmate        <NA>          <NA>
## 141         16416       3         Playmate        <NA>          <NA>
## 142         16483       3     Investigator        <NA>          <NA>
## 143         16507       3         Playmate        <NA>          <NA>
## 144         16516       3           Mother        <NA>        Female
## 145         16523       3           Mother        <NA>        Female
## 146         16535       3     Investigator        <NA>          <NA>
## 147         16536       3         Playmate        <NA>          <NA>
## 148         16560       3         Playmate        <NA>          <NA>
## 149         16563       3      Grandmother        <NA>        Female
## 150         16580       3         Playmate        <NA>          <NA>
## 151         16587       3           Mother        <NA>        Female
## 152         16603       3     Investigator        <NA>          <NA>
## 153         16630       3         Playmate        <NA>          <NA>
## 154         16644       3     Investigator        <NA>          <NA>
## 155         16658       3           Cousin       5;1.0          <NA>
## 156         16662       3            Adult        <NA>          <NA>
## 157         16680       3     Investigator        <NA>          <NA>
## 158         16682       3           Mother        <NA>        Female
```

```
## 159         16684       3         Playmate        <NA>          <NA>
## 160         16704       3     Investigator        <NA>          <NA>
## 161         16729       3           Cousin      17;0.0          <NA>
## 162         16764       3         Playmate        <NA>          <NA>
## 163         16446       2     Investigator        <NA>          <NA>
## 164         16447       2           Mother        <NA>        Female
## 165         16463       2     Investigator        <NA>          <NA>
## 166         16466       2            Uncle        <NA>          Male
## 167         16474       2      Grandmother        <NA>        Female
## 168         16490       2     Investigator        <NA>          <NA>
## 169         16492       2           Mother        <NA>        Female
## 170         16510       2         Playmate        <NA>          <NA>
## 171         16522       2     Investigator        <NA>          <NA>
## 172         16547       2            Uncle        <NA>          Male
## 173         16565       2     Investigator        <NA>          <NA>
## 174         16572       2     Investigator        <NA>          <NA>
## 175         16614       2     Investigator        <NA>          <NA>
## 176         16648       2     Investigator        <NA>          <NA>
## 177         16665       2     Investigator        <NA>          <NA>
## 178         16706       2           Cousin       5;4.0          <NA>
## 179         16716       2         Playmate        <NA>          <NA>
## 180         16718       2            Adult        <NA>          <NA>
## 181         16726       2     Investigator        <NA>          <NA>
## 182         16743       2          Brother       8;0.0          Male
## 183         16758       2            Adult        <NA>          <NA>
## 184         16765       2           Mother        <NA>        Female
## 185         16766       2           Sister      10;0.0        Female
## 186         16393       1      Grandmother        <NA>        Female
## 187         16404       1         Playmate        <NA>          <NA>
## 188         16413       1           Mother        <NA>        Female
## 189         16417       1            Uncle        <NA>          Male
## 190         16419       1      Grandmother        <NA>        Female
## 191         16436       1            Adult        <NA>          <NA>
## 192         16442       1            Adult        <NA>          <NA>
## 193         16452       1           Mother        <NA>        Female
## 194         16493       1           Cousin       8;3.0          <NA>
## 195         16504       1      Grandmother        <NA>        Female
## 196         16506       1     Investigator        <NA>          <NA>
## 197         16515       1     Investigator        <NA>          <NA>
## 198         16530       1           Cousin       5;3.0          <NA>
## 199         16544       1           Mother        <NA>        Female
## 200         16556       1           Mother        <NA>        Female
## 201         16566       1         Playmate        <NA>          <NA>
## 202         16570       1      Grandmother        <NA>        Female
## 203         16579       1     Investigator        <NA>          <NA>
## 204         16596       1         Playmate        <NA>          <NA>
## 205         16598       1         Playmate        <NA>          <NA>
## 206         16615       1           Mother        <NA>        Female
## 207         16616       1            Adult        <NA>          <NA>
## 208         16618       1         Playmate        <NA>          <NA>
## 209         16622       1           Mother        <NA>        Female
## 210         16628       1         Teenager      15;0.0          <NA>
## 211         16636       1           Mother        <NA>        Female
```

```
## 212         16637       1           Cousin       8;8.0          <NA>
## 213         16638       1            Adult        <NA>          <NA>
## 214         16672       1            Adult        <NA>          <NA>
## 215         16692       1     Investigator        <NA>          <NA>
## 216         16696       1     Investigator        <NA>          <NA>
## 217         16701       1         Playmate        <NA>          <NA>
## 218         16723       1         Playmate        <NA>          <NA>
## 219         16728       1          Brother       4;5.0          Male
## 220         16752       1     Investigator        <NA>          <NA>
## 221         16759       1         Playmate        <NA>          <NA>
## 222         16760       1     Investigator        <NA>          <NA>
```

```
#same for Wells
wells<-read.csv(file="/Users/lscpuser/Downloads/Wells/output2.csv", header=FALSE, sep=
";")
names(wells)<-c("speaker", "utterance", "directed")
wells_speakers<- wells %>% group_by(speaker) %>% summarise(no_rows = length(speaker))
wells_speakers<- wells_speakers %>%   arrange(desc(no_rows))
wells_speakers
```

```
## # A tibble: 136 x 2
##    speaker no_rows
##    <fct>     <int>
##  1 MOT       29718
##  2 FAT        3943
##  3 GRA        1931
##  4 NIC        1627
##  5 UNK        1549
##  6 TRA        1018
##  7 ROB        1015
##  8 MAT        1009
##  9 RIC         808
## 10 RAC         672
## # … with 126 more rows
```

```
#6 CHECK : %dir annotation names
wells_directed<- wells %>% group_by(directed) %>% summarise(no_rows = length(directed))
wells_directed<- wells_directed %>% as_tibble %>%   arrange(desc(no_rows))
wells_directed<- wells_directed %>%   arrange(desc(no_rows))
wells_directed
```

```
## # A tibble: 210 x 2
##    directed  no_rows
##    <fct>       <int>
##  1 ""          52230
##  2 %add to M     191
##  3 %add to D      86
##  4 %add to E      71
##  5 %add to T      66
##  6 %add to R      56
##  7 %add to F      45
##  8 %add to C      39
##  9 %add to H      36
## 10 %add to S      36
## # … with 200 more rows
```

Utterances per speaker Demuth corpus : see sesotho_speakers Utterances per speaker Demuth corpus with directed annotation : see sesotho_directed_si

Utterances per speaker Wells corpus : see wells_speakers Utterances per speaker Wells corpus with directed annotation : see wells_directed