Appendix



Figure A1. Boxplot of distribution of intensity differences for all children from youngest (left) to oldest (right) by birthplace.



Figure A2. Scatterplot of correlation between intensity differences and FRCNHV by birthplace with 99% confidence intervals.

*R code to calculate FRC for segmental contexts from a reference corpus as a .txt file*

# Script to calculate FRCs in reference corpus

# NOTE -- corpus should be a .txt file and punctuation should be replaced by something like PUN unless you need to do so otherwise.

# Packages you'll need to load

library(stringr)

library(ngram)

library(dplyr)

# Load the corpus and get it to a bigram table ####

# Select corpus from your computer

corpus <- scan(file = file.choose(), what = "character", sep = " ", comment.char = ".", encoding = "UTF-8", quote = "")

head(corpus)

# Get rid of empty characters

corpus <- corpus[nchar(corpus)> 0]

head(corpus)

# Make upper case

corpus <- toupper(corpus)

head(corpus)

#print(corpus)

# make the corpus into a string, remove commas, and then make a bigram table

new.corpus <- toString(corpus)

new.corpus <- gsub(",","",toString(new.corpus))

bigrams <- ngram(new.corpus, n =2)

bigrams\_df <- get.phrasetable(bigrams) # there is a space after every pair.

# get rid of space

bigrams\_df$ngrams <- as.character(bigrams\_df$ngrams)

bigrams\_df$ngrams <- trimws(bigrams\_df$ngrams)

# edit names of columns in bigrams\_df

names(bigrams\_df)[names(bigrams\_df) == "ngrams"] <- "bigram"

names(bigrams\_df)[names(bigrams\_df) == "freq"] <- "bigram\_freq"

# separate bigrams into word 1 and word 2 columns

bigrams\_df[c('word\_1', 'word\_2')] <- str\_split\_fixed(bigrams\_df$bigram, ' ', 2)

# Manual step - clean bigram file and manually subset as necessary #####

# save table

write.table(bigrams\_df, file = "YOURDIRECTORY/bigramfilename.txt", quote = FALSE, sep = "\t", row.names = FALSE)

# Calculate FRC for each word (use for all corpora except Davies) ####

# load cleaned bigram table

bigrams\_df <- read.delim("YOURDIRECTORY/bigramfilename\_edited.txt", stringsAsFactors = TRUE)

# create another column to code the preceding context

# NOTE - whatever you do in the prec\_context column should help you calculate the FRC values later on.

bigrams\_df$prec\_context <- bigrams\_df$word\_1

# limit word\_1 to the last letter

# NOTE - this works in my case, but you may need something else in this column.

bigrams\_df$prec\_context <- str\_sub(bigrams\_df$prec\_context,-1)

# create another column for coding for prec context

# NOTE - you'll use context\_type to calculate the FRC values

bigrams\_df$context\_type <- as.character(bigrams\_df$prec\_context)

# Examples of how to code context\_type for the FRC calcs

# For FRC based on post- #, l, m, n

#bigrams\_df$context\_type[bigrams\_df$context\_type == "N"] <- "non-reducing"

#bigrams\_df$context\_type[bigrams\_df$context\_type == "M"] <- "non-reducing"

#bigrams\_df$context\_type[bigrams\_df$context\_type == "L" & startsWith(as.character(bigrams\_df$word\_2), "D")] <- "non-reducing"

#bigrams\_df$context\_type[bigrams\_df$context\_type != "non-reducing"] <- "reducing"

# For FRC based on NHVs

# standardize accent

# bigrams\_df$context\_type[bigrams\_df$context\_type == "Í"] <- "I"

# bigrams\_df$context\_type[bigrams\_df$context\_type == "É"] <- "E"

# bigrams\_df$context\_type[bigrams\_df$context\_type == "Á"] <- "A"

# bigrams\_df$context\_type[bigrams\_df$context\_type == "Ó"] <- "O"

# bigrams\_df$context\_type[bigrams\_df$context\_type == "Ú"] <- "U"

# code for context type

# bigrams\_df$context\_type[bigrams\_df$context\_type == "E"] <- "reducing"

# bigrams\_df$context\_type[bigrams\_df$context\_type == "A"] <- "reducing"

# bigrams\_df$context\_type[bigrams\_df$context\_type == "O"] <- "reducing"

# bigrams\_df$context\_type[bigrams\_df$context\_type != "reducing"] <- "non-reducing"

# filter out NA values

bigrams\_df <- bigrams\_df %>%

filter(!is.na(bigram\_freq)) %>%

filter(!is.na(context\_type))

# make data frame to hold values

frc\_values <- data.frame(matrix(ncol = 5, nrow = 1))

colnames(frc\_values) <- c("word","lex\_freq","num\_non\_reducing", "num\_reducing", "FRC")

# run through loop to add FRCs to bdg\_frc\_ADS\_values

bigrams\_df$word\_2 <- factor(bigrams\_df$word\_2)

words <- levels(bigrams\_df$word\_2)

for (i in 1:length(words)) {

freq\_value <- sum(bigrams\_df$bigram\_freq[bigrams\_df$word\_2 == words[i]])

num\_non\_reducing <- sum(bigrams\_df$bigram\_freq[bigrams\_df$word\_2 == words[i] & bigrams\_df$context\_type == "non-reducing"])

num\_reducing <- sum(bigrams\_df$bigram\_freq[bigrams\_df$word\_2 == words[i] & bigrams\_df$context\_type == "reducing"])

frc\_value <- num\_reducing/freq\_value

frc\_values[nrow(frc\_values) + 1,1:5] = c(words[i],freq\_value,num\_non\_reducing,num\_reducing,frc\_value)

rm(freq\_value,num\_non\_reducing,num\_reducing,frc\_value)

}

# done!

# remove first row

frc\_values <- frc\_values[-1,]

# correct vector types

frc\_values$word <- as.factor(frc\_values$word)

frc\_values$lex\_freq <- as.numeric(frc\_values$lex\_freq)

frc\_values$num\_non\_reducing <- as.numeric(frc\_values$num\_non\_reducing)

frc\_values$num\_reducing <- as.numeric(frc\_values$num\_reducing)

frc\_values$FRC <- as.numeric(frc\_values$FRC)

# save your values!!

write.table(frc\_values, file = "YOURDIRECTORY/frcs.txt", quote = FALSE, sep = "\t", row.names = FALSE)

Table A1. *Mixed-effects linear regression predicting intensity differences 5- to 8-year-olds’ (N = 15) word-initial /d/; data excludes* de *and* después *(*n *= 348)*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | *β* | *SE* | *t*-value | *p*-value | *n* observations | mean ID |
| Intercept  (nasal, lateral, pause; stressed) | 15.43 | 1.57 | 9.80 | < .001 |  |  |
| Preceding context  non-high vowel  other V/C | -4.72  -1.32 | 0.85  0.95 | -5.46  -1.39 | < .001  .16 | 309  134 | 6.60  8.67 |
| Lexical stress | -2.81 | 0.79 | -3.52 | < .001 | 319 | 6.97 |
| Log(Lexical frequency) | -0.31 | 0.22 | -1.47 | .15 |  |  |
| Scale(FRC) | 5.55 | 1.97 | 2.82 | .005 |  |  |
| Preceding context:Scale(FRC)  non-high vowel  other V/C | -3.42  -3.82 | 1.71  1.80 | -2.00  -2.12 | .04  .03 | 309  134 | 6.60  8.67 |
| Log(Lexical frequency):Scale(FRC) | -0.77 | 0.34 | -2.32 | .02 |  |  |
| Random intercepts: Word: σ2 = 1.97; Speaker: σ2 = 2.76; Residual: σ2 = 24.46  Model fit: AIC = 2141.19; BIC = 2187.41; conditional R2 = 0.35 | | | | | | |

The word *después* was removed from this model since all the tokens of this word came from one child who produced this word in the phrase *y* *después* and did so consistently with very little constriction. The word *de* was excluded to demonstrate that the interaction between lexical frequency and FRC holds without this extremely frequent word.

Table A2. *Mixed-effects linear regression predicting intensity differences 9 to 15-year-olds’ (N = 14) word-initial /d/; data excludes* de *and* después *(*n *= 244)*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | *β* | *SE* | *t*-value | *p*-value | *n* observations | mean ID |
| Intercept  (nasal, lateral, pause) | 12.03 | 1.35 | 8.88 | < .001 |  |  |
| Preceding context  non-high vowel  other V/C | -6.52  -4.41 | 0.85  0.93 | -7.59  -4.71 | < .001  < .001 | 246  141 | 4.62  6.76 |
| Log(Lexical frequency) | -0.12 | 0.16 | -0.77 | .44 |  |  |
| Scale(FRC) | 5.12 | 1.74 | 2.93 | .003 |  |  |
| Preceding context:Scale(FRC)  non-high vowel  other V/C | -3.51  -2.87 | 1.67  1.73 | -2.09  -1.65 | .03  .09 | 246  141 | 4.62  6.76 |
| Log(Lexical frequency):Scale(FRC) | -0.65 | 0.25 | -2.63 | .009 |  |  |
| Random intercept: Speaker: σ2 = 2.08; Residual: σ2 = 17.22  Model fit: AIC = 1407.68; BIC = 1442.65; conditional R2 = 0.35 | | | | | | |

The words *de* and *después* were removed from this model to match the model for the younger children. When these words are included, the results are the same with the exception that speech rate is also included in the model.