28-4-2015

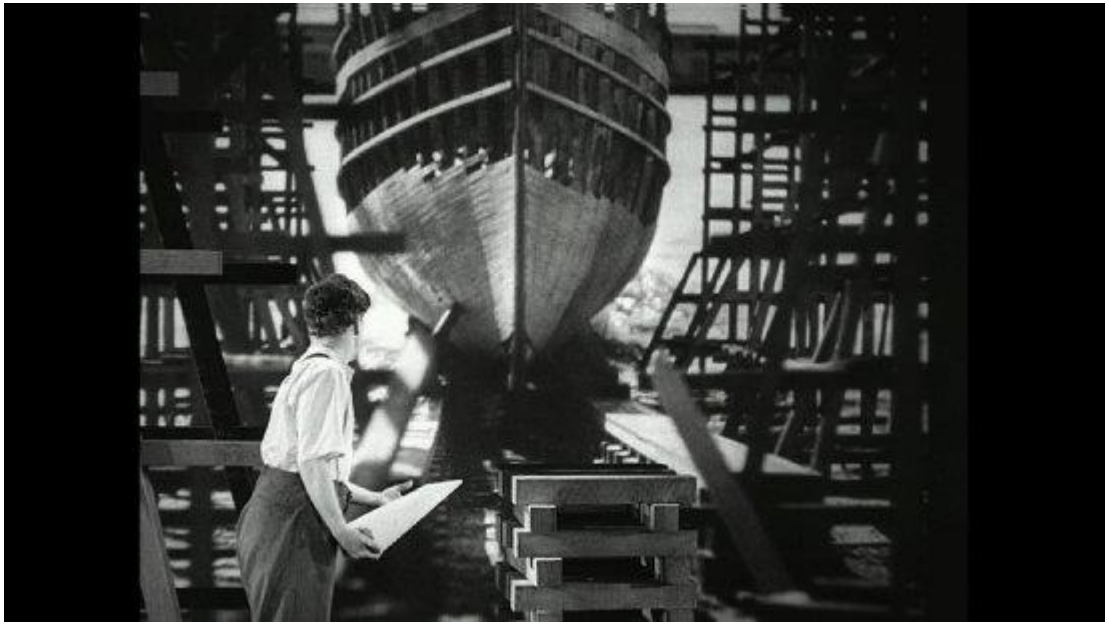
Predictors of disfluency markers in first language attrition

Ingemarie Donker

First of all:

Many thanks to Bregtje Seton for the data!!

Good Ol’ Charlie



The plan for the coming 30 minutes

- L1 Attrition & Disfluencies
- Some Variables
- The Experiments
- The Data
- How to go about analyzing it
- Finding The Right Model
- Finally, The Results
- What we may conclude

L1 Attrition

- During the acquisition of another language, the native language is not necessarily stable as it is perhaps being used less and may also be influenced by the addition of another language.

- This process is called *first language attrition* and is especially true for someone who changes their language environment and becomes dominant in a second language (L2).

Disfluency

- L1 attrition can manifest itself in different forms, but it is assumed that it affects lexical access more than other aspects of language (Schmid & Jarvis, 2014; Schmid & Köpke, 2009)

- If the language becomes less accessible, this could be a cause for disfluent speech, because it takes more time to plan and produce.

Disfluency Markers

› Disfluency types as categorized by Levelt (1983)

- Appropriateness Repairs: per 1000 words (A1000)
- Error repairs: per 1000 words (E1000)
- Difference Repairs: per 1000 words (DR1000)
- Isolated Filled Pauses: per 1000 words (IFP1000)
- Repetitions: per 1000 words (R1000)

Research Questions

- Is there a difference between Attriters and Native Speakers (NSs) on how many disfluencies they have in their spontaneous speech?
- If so, are the factors of AoE, LoR, attitude, L1 use, working memory, and L1 proficiency relevant predictors that play a role in this process?
- In general, is there an influence of the number of errors people make on the number of disfluencies, and is there an influence of lexical diversity on the number of disfluencies?

Hypotheses

- **AoE**: older when migrating = less disfluencies
- **LoR**: longer stay = more disfluencies
- **L1 Use**: more use = less disfluencies
- **L1 Motivation**: more positive = less disfluencies
- **L1 proficiency**: more proficient = less disfluencies
- **Errors**: more errors = more disfluencies
- **Lexical Diversity**: more diverse = less disfluencies

Background Variables (1)

General & Working Memory

- **Age**: age at testing
- **AoE**: Age of emigration (only for attriters)
- **LoR:** Length of residence (only for attriters)
- **Gender**
- **Location**: location of testing (Chicago (CH), London (LD), Toronto (TO), Groningen (GR), or Leiden (LE))
- **N2backdp**: the dprime score on a working memory nback task

Background Variables (2)

Proficiency & Use

- **HolProf**: the holistic proficiency of the participant as rated by three independent raters (icc interrater reliability of 0.91) - total score out of 90
- **DuCtest**: score on Dutch C-test - percentage
- **EnCtest**: score on English C-test (only for attriters) - percentage
- **Use**: average self-reported use of Dutch at home + work + elsewhere - percentage

Background Variables (3)

Motivation & Errors & Lexical Diversity

- **MotivationTotal**: motivation to keep speaking Dutch + cling to culture+ return to Holland - percentage
- **TotE1000**: Total number of errors per 1000 words
- **Guiraud**: Number of Types divided by the Square Root of the Tokens: measure of lexical diversity.
- **VOCD**: the VOCD type token ratio as a measure of lexical diversity in the retellings

Participants

	Controls (n=27)	Attriters (n=54)
Age	47 (18 – 68)	47 (19-69)
Age of Emigration	-	24 (5-42)
Length of Residence	-	28 (5-56)

Methodology

› Free speech data were elicited by the Charlie Chaplin film retelling task (Schmid & Beers Fägersten, 2010)

› Orthographic transcription in CHAT format

› Coding of different types of disfluency markers according to CHILDES coding standards

› General linear mixed-effects model using the package glmmADMB in R (data treated according to negative binomial distribution)

Why Mixed-Effects?

- The method is relatively easy and does not require a balanced design
- Mixed-effects models are robust to missing data
- Difference between fixed-effect (e.g. word category) and random-effect (e.g. subject) factors
- Makes the regression formula as precise as possible for every individual observation in our random effects, so allows specific models for every observation and for every subject

Specific Model

- *Fixed-effects*: count data on the different disfluencies, with every participants having five different scores on the different disfluency types
- → plus all the other predictor variables

- *Random-effects*: participant
- → tested for random slopes and random intercepts of participant and group with Disfluency Type

Procedure/Choices

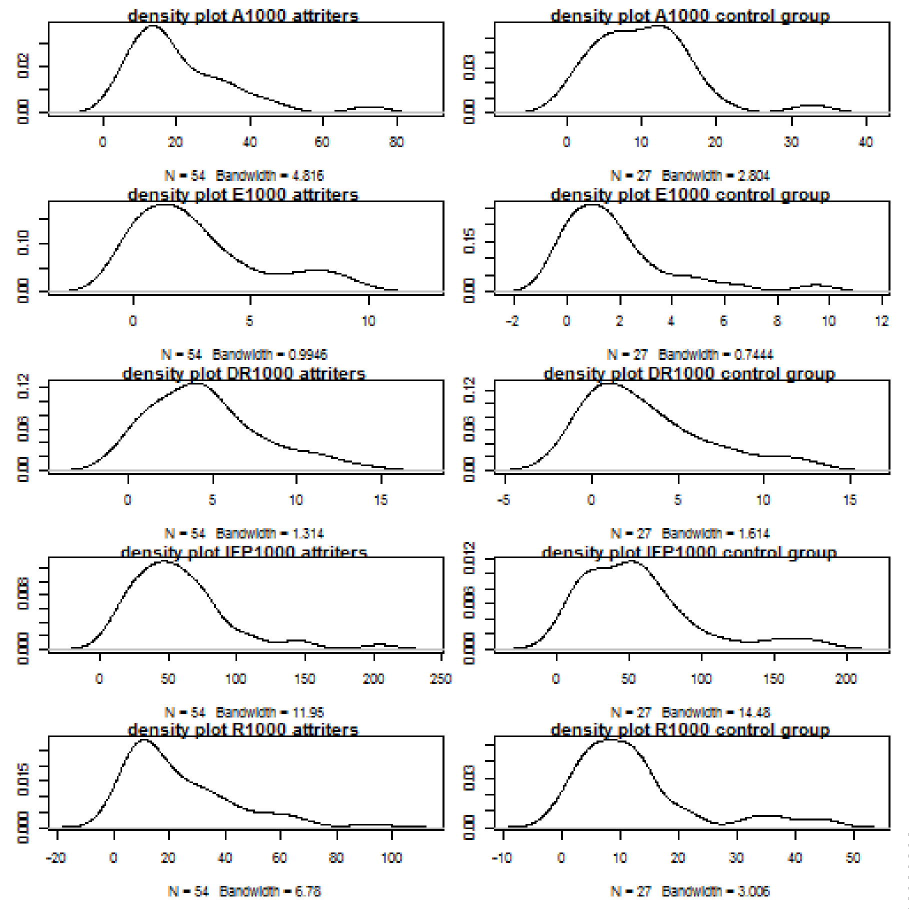
- Summary of the background variables of the native control group and the attriters (plots, tables, histograms, boxplots, density, Wilcoxon rank sum tests, Shapiro-Wilk normality tests)
- Center numerical predictors
- Check correlating predictor variables (Spearman)

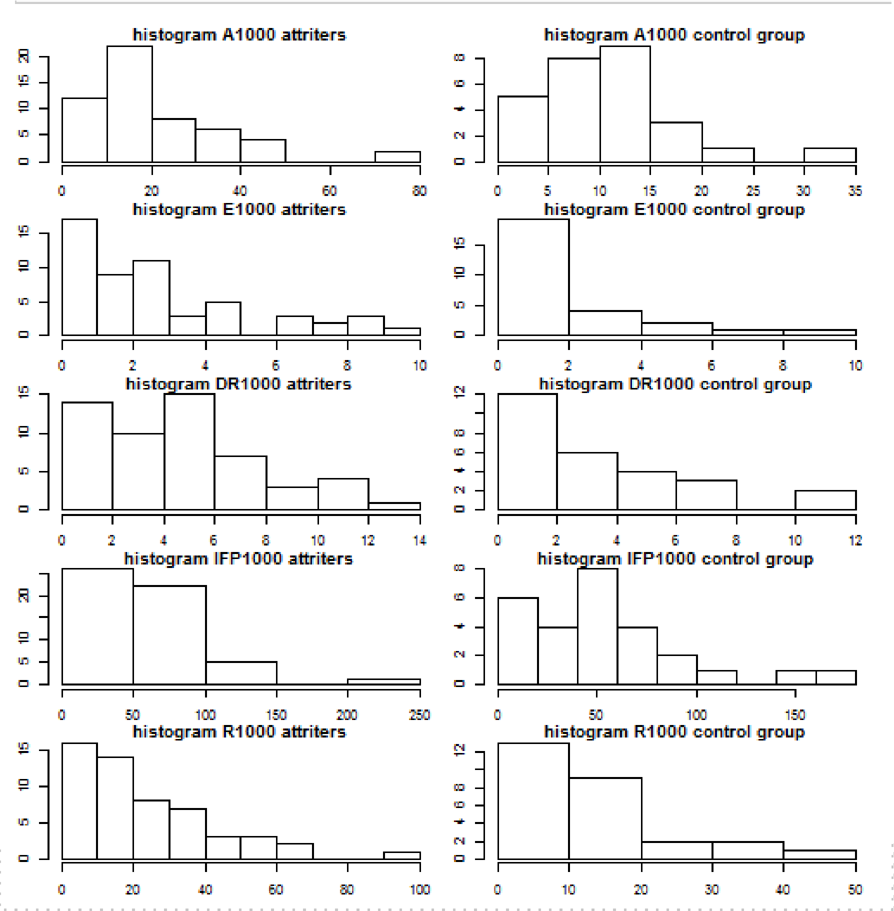
→ age correlates strongly with LoR

Normal Distribution

- The different disfluency measures are all skewed to the right.
- Some measures are good with a simple log transformation, but the ones that contain zero's get a bimodal distibution with a log1p.

Visually

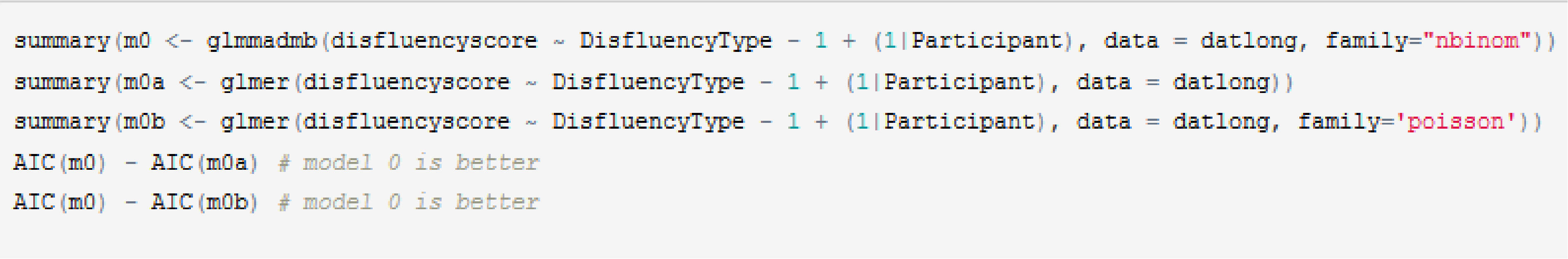

density plot A1000 attriters
0.00 0.02
0 20 40 60 80
N = 54 Bandwidth = 4.816
density plot A1000 control group
0.00 0.03
0 10 20 30 40
N = 27 Bandwidth = 2.804
density plot E1000 attriters
0.00 0.10
0 5 10
N = 54 Bandwidth = 0.9946
density plot E1000 control group
0.00 0.15
-2 0 2 4 6 8 10 12
N = 27 Bandwidth = 0.7444
density plot DR1000 attriters
0.00 0.06 0.12
0 5 10 15
N = 54 Bandwidth = 1.314
density plot DR1000 control group
0.00 0.06 0.12
-5 0 5 10 15
N = 27 Bandwidth = 1.614
density plot IFP1000 attriters
0.000 0.006
0 50 100 150 200 250
N = 54 Bandwidth = 11.95
density plot IFP1000 control group
0.000 0.006 0.012
0 50 100 150 200
N = 27 Bandwidth = 14.48
density plot R1000 attriters
0.000 0.015
-20 0 20 40 60 80 100
N = 54 Bandwidth = 6.78
density plot R1000 control group
0.00 0.03
-10 0 10 20 30 40 50
N = 27 Bandwidth = 3.006


histogram A1000 attriters
0 5 10 15 20
0 20 40 60 80
histogram A1000 control group
0 2 4 6 8
0 5 10 15 20 25 30 35
histogram E1000 attriters
0 5 10 15
0 2 4 6 8 10
histogram E1000 control group
0 5 10 15
0 2 4 6 8 10
histogram DR1000 attriters
0 5 10 15
0 2 4 6 8 10 12 14
histogram DR1000 control group
0 2 4 6 8 12
0 2 4 6 8 10 12
histogram IFP1000 attriters
0 5 10 20
0 50 100 150 200 250
histogram IFP1000 control group
0 2 4 6 8
0 50 100 150
histogram R1000 attriters
0 5 10 15
0 20 40 60 80 100
histogram R1000 control group
0 2 4 6 8 12
0 10 20 30 40 50

Poisson vs Negative Binomials

- Most of these variables show a distribution that is more similar to a Poisson distribution.
- Count data where the **maximum possible counts is unknown**, so do glmers with Poisson or Negative Binomials, where each participant has 5 different count values for each different Disfluency Type.

Poisson vs Binomial

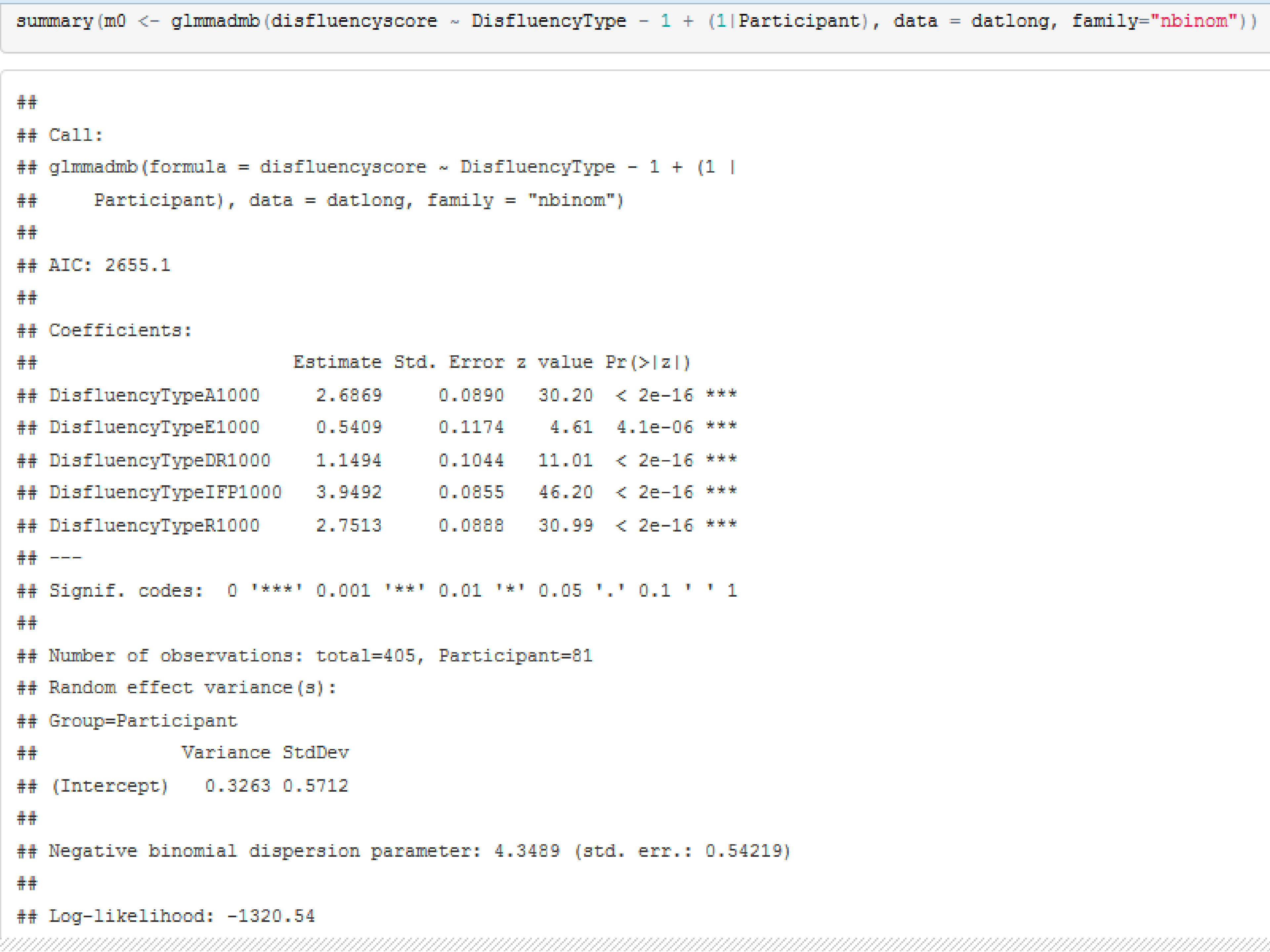


```
summary(m0 <- glmmadmb(disfluencyscore ~ DisfluencyType - 1 + (1|Participant), data = datlong, family="nbinom"))
summary(m0a <- glmer(disfluencyscore ~ DisfluencyType - 1 + (1|Participant), data = datlong))
summary(m0b <- glmer(disfluencyscore ~ DisfluencyType - 1 + (1|Participant), data = datlong, family='poisson'))
AIC(m0) - AIC(m0a) # model 0 is better
AIC(m0) - AIC(m0b) # model 0 is better
```

Because there is quite some overdispersion, a Negative Binomial Model is better than Poisson

Model building (1)

- **Group** and **Disfluency Type** were added to the model first and at this point there was a significant effect of group, the attriters having more disfluencies than the native speaker control group



```
summary(m0 <- glmmadmb(disfluencyscore ~ DisfluencyType - 1 + (1|Participant), data = datlong, family="nbinom"))
```

```
##
## Call:
## glmmadmb(formula = disfluencyscore ~ DisfluencyType - 1 + (1 |
##     Participant), data = datlong, family = "nbinom")
##
## AIC: 2655.1
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## DisfluencyTypeA1000     2.6869     0.0890   30.20  < 2e-16 ***
## DisfluencyTypeE1000     0.5409     0.1174    4.61  4.1e-06 ***
## DisfluencyTypeDR1000    1.1494     0.1044   11.01  < 2e-16 ***
## DisfluencyTypeIFP1000   3.9492     0.0855   46.20  < 2e-16 ***
## DisfluencyTypeR1000     2.7513     0.0888   30.99  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of observations: total=405, Participant=81
## Random effect variance(s):
## Group=Participant
##             Variance StdDev
## (Intercept)   0.3263 0.5712
##
## Negative binomial dispersion parameter: 4.3489 (std. err.: 0.54219)
##
## Log-likelihood: -1320.54
```

Model Building (2)

- Add variables to the statistical model:

- Attriter-specific measures:
  - Age of Emigration (AoE)
  - Length of Residence (LoR)
  - Attitude/Motivation
  - L1 Use.

Results (1)

- Difference between attriters and NSs disappears after other factors included, so no extra group difference
- The attriter-specific variables: not a very strong effect on the number of disfluencies:
  - **LoR**: higher LoR = more error repairs
  - Interaction **Motivation** with **AoE**:

  → lower AoE = higher role of motivation in number disfluencies

  → here, higher motivation = less disfluencies

Model Building (3)

› Other variables of interest:
  - L1 Holistic proficiency
  - C-test scores
  - Total number of errors (accuracy)
  - Guiraud (complexity)
  - Working Memory

Results (2)

- **L1 Holistic proficiency** *(only predictor of number of disfluencies for NSs, not attriters)*

  → higher rating HP = less disfluencies
- **C-test**: higher C-test score = more disfluencies
- **Complexity** *(only significant for NSs):*

  → higher lexical diversity = less disfluencies
- **Accuracy**: more errors = more disfluencies

Model Building

- Add predictors and compare models, if t>2 then the new model is significantly better
- Number of models compared: too many to count

Final model:

- summary(m3e <- glmmadmb(disfluencyscore ~ DisfluencyType + DisfluencyType:Group + IsAttriters:cMotivationTotal:cAoE + IsAttriters:cLoR:IsE1000 + cTotE1000 + cHolProf:IsNatives + cGuiraud:IsNatives + cDuCtest - 1 + (1|Participant), data=datlong, family='nbinom'))

Conclusion (1)

- Attriter-specific factors did not come out as very strong predictors of disfluent speech, however:

- **LoR** seemed to influence the number of error repairs
- **Attitude** and **motivation** seemed to play a role for younger attriters on the disfluencies in general.

Conclusion (2)

- **Holistic proficiency** only has an effect for the native control group
- **C-test scores** have a positive effect on the disfluencies
- **Accuracy** (total number of errors) is also significantly related to the number of disfluencies
- **Complexity** only has an effect for NSs

General conclusion

- Absence of a link between the number of disfluencies in the attriters and the more lexically diverse vocabulary use or in this group signifies a clear problem of lexical access and **not** of language loss .
- The attriters are able to eventually access their lexical items, but in order to do this they suffer more from disfluent speech.

Conclusion MM

- A complicated design with many variables can best be modeled with a mixed effects regression model, which can give more information about the combined origin of the disfluencies
- Disfluencies and the background variables not treated by themselves and therefore do not cause a problem of multiple comparisons

Thank you!

Some References

- Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition, 14*, 41-104.
- Schmid, M. S., & Jarvis, S. (2014). Lexical first language attrition. *Bilingualism: Language and Cognition*, *17*(4), 729–748.
- Schmid, M. S., & Köpke, B. (2009). L1 attrition and the mental lexicon. In A. Pavlenko (Ed.), *The Bilingual Mental Lexicon: Interdisciplinary Approaches* (pp. 209–238). Clevedon: Multilingual Matters.

Discussion

- Overdisperison: Poisson or Negative Binomial Model?

- Two models are just as good, but one has an extra significant predictor: okay to choose that one?

- Higher C-test score = more disfluencies
  → what does this test measure?

```
›  Coefficients:
›                                           Estimate Std. Error  z value   Pr(>|z|)
›  DisfluencyTypeA1000                      2.597196  0.229590   11.31     < 2e-16 ***
›  DisfluencyTypeE1000                      0.685137  0.272090    2.52      0.0118 *
›  DisfluencyTypeDR1000                     1.250735  0.249990    5.00      5.6e-07 ***
›  DisfluencyTypeIFP1000                    4.294344  0.225510   19.04     < 2e-16 ***
›  DisfluencyTypeR1000                      2.733485  0.226400   12.07     < 2e-16 ***
›  cTotE1000                                0.028438  0.007640    3.72      0.0002 ***
›  cDuCtest                                 0.012968  0.005538    2.34      0.0192 *
›  DisfluencyTypeA1000:Groupattriters       0.275176  0.249580    1.10      0.2702
›  DisfluencyTypeE1000:Groupattriters      -0.104312  0.303720   -0.34      0.7313
›  DisfluencyTypeDR1000:Groupattriters      0.010056  0.275510    0.04      0.9709
›  DisfluencyTypeIFP1000:Groupattriters    -0.369180  0.243770   -1.51      0.1299
›  DisfluencyTypeR1000:Groupattriters       0.178147  0.246270    0.72      0.4694
›  cHolProf:IsNatives                      -0.052636  0.018639   -2.82      0.0047 **
›  IsNatives:cGuiraud                      -0.361911  0.124570   -2.91      0.0037 **
›  IsAttriters:cMotivationTotal:cAoE        0.000772  0.000379    2.04      0.0415 *
›  IsAttriters:cLoR:IsE1000                 0.021141  0.007758    2.73      0.0064 **
```