Assignment 1 - Language Development in ASD - part 2

Riccardo Fusaroli

July 7, 2017

# Language development in Autism Spectrum Disorder (ASD)

Background: Autism Spectrum Disorder is often related to language impairment. However, this phenomenon has not been empirically traced in detail: i) relying on actual naturalistic language production, ii) over extended periods of time.

We therefore videotaped circa 30 kids with ASD and circa 30 comparison kids (matched by linguistic performance at visit 1) for ca. 30 minutes of naturalistic interactions with a parent. We repeated the data collection 6 times per kid, with 4 months between each visit. We transcribed the data and counted: i) the amount of words that each kid uses in each video. Same for the parent. ii) the amount of unique words that each kid uses in each video. Same for the parent. iii) the amount of morphemes per utterance (Mean Length of Utterance) displayed by each child in each video. Same for the parent.

This data is in the file you prepared in the previous class.

NB. A few children have been excluded from your datasets. We will be using them next week to evaluate how good your models are in assessing the linguistic development in new participants.

We then want to test the language trajectory of child and parent over time.

This RMarkdown file is structured in the following way:

1. The exercises: read them carefully. Under each exercise you will have to write your answers, once you have written and run the code. This is the part that you have to directly send to the teachers.
2. An (optional) guided template full of hints for writing the code to solve the exercises. Fill in the code and the paragraphs as required. Then report your results under the exercise part.
3. In exercise 4 you will be asked to create the best possible model of language development in TD and ASD children, picking and choosing whatever additional variables you want from the dataset. Next time, the models produced by the different groups will compete against each other to see who can produce the best model, so choose carefully!

You will have to have a github repository for the code and send the answers to Malte and Riccardo without code (but a link to your github/gitlab repository). This way we can check your code, but you are also forced to figure out how to report your analyses :-)

Remember to submit only your findings, and not just the code. To do this you can either - Write your answers in a separate document - Write your answers in the template, but tell rstudio not to print the code chunks when you knit it with the chunk option include=FALSE

## [1] 11

To re-iterate one more time: Hand in a document with your findings but without code (html or pdf or word) and a link to your github/gitlab with your Rmd file.

N.B. The following lines are a summary of the questions to be answered, the step-by-step instructions and tips are in the template.

## Exercise 1) Preliminary Data Exploration

Describe the participant samples in the dataset (e.g. by diagnosis, age, etc.). Do you think the two groups are well balanced? If not, what do you think was the reason?

Participants samples are well balanced according to - Diagnosis - almost the same number of participants, 10 more TD participants - Age - Types\_MOT - spread all over but are slightly different, ASD group peaks at 350 and TD group peaks at 400

not well balanced - MLU\_MOT - mothers of TD children has slightly higher mean than ASD group - CHI\_MLU - TD group says much more words while ASD group says the most from 1-1,5 - types\_CHI - ASD group says much less(up to 50) unique words in comparison to TD group (up to 200) - token\_CHI - TD group said much more words (up to 800) while ASD up to 200

### Exercise 2) Children learning language: the effects of time and ASD

Describe linguistic development in TD and ASD children in terms of Mean Length of Utterance (MLU)?

In our model, we used child’s MLU as an a dependent variable, visit and diagnosis as independent variables. We also included random effects, specifically subject as random intercept and visit as random slope. After analysis, we found the visit is a significant predictor of child’s MLU , β = 0.23,(SE = 0.02), t = 9.516, p < .0001. However, diagnosis did not turn out significant, β = 0.29 ,(SE = 0.15), t = 1.91, p > .05. The result indicates that ASD children’s linguistic performance is not different from non-ASD children. But it also suggests that overall children MLU is changing with each visit. Aditionally, we calcualted r2 for our model, which was following: R2m=0.219 and R2c = 0.803. This indicates that our model with random effects explains significatly more variance but we cannot further generalize it.

### Exercise 3) Child directed speech as a moving target

Describe how parental use of language changes over time in terms of MLU. What do you think is going on?

According to our analysis, mothers’ MLU changes depending on Diagnosis, β = 0.50,(SE = 0.11), t = 4.42, p < .0001. And diagnosis is a significant predictor as well as time (in our case Visit), β = 0.12, (SE = 0.02), t = 6.60, p < .0001. This might suggest that knowing of diagnosis changes mothers’ speech accordingly. Moreover, the R2 for the model: R2m = 0.230 and R2c= 0.676, similarly suggests that including of random effects helps to explain more variance but cannot be generalized.

### Exercise 4) Looking into “individual differences” (demographic, clinical or cognitive profiles)

The dataset contains some additional variables characterizing the kids’ cognitive and clinical profile: ADOS (autism severity), MSEL EL (Expressive Language, that is, verbal IQ, or linguistic skills at first visit as assessed by a psychologist using Mullen Scales of Early Learning), MSEL VR (Visual Reception, used as a proxy for non verbal IQ at first visit), Age, Gender, Ethnicity. Would it make sense to add any of them to your model of linguistic trajectories? Create the best possible model (the one that best explain the data, with MLU as outcome). Next time your model will be tested on new participants, and we will proclaim a winner. Describe your strategy to select the best models (how did you choose the variables to include?) and send the code to Riccardo and Celine.

In order to identify best model, that describes our data, we have chosen as independant variable VISIT to predict Child MLU because we are interested in linguitic trajectory. We belive that we can’t use additional clinical variables (e.g. noverbalIQ or ADOS), because they were collected not every visit, thus we don’t have enough data. Additionally, we see no difference in variable AGE and Visit, so it makes no sence to add Age as IV. In our opinion, ethnicity and gender plays no role. We believe that child MLU is calculated from overall amount of words, which is token\_CHI variable. Therefore it is not reasonable to make token\_CHI as predictor. We did correlation analysis between CHI\_MLU and types\_chi and we discovered strong correlation. In our opinion trying to predict one index of liguistic performance, which is Child MLU, by another index of linguistic performance doesn’t make much sense, because they belong to one category and have strong correlation which violates assumption of collinearity. Also if we do so, our main predictor, which is VISIT becomes not significant. However, our main priority is to see linguistic trajectory over time. Originally the idea was that Diagnosis could be a significant predictor and in model where we included only Diagnosis as a predictor without visit a slope, we found that diagnosis is significant predictor but in the model that predicts MLU from both Visit and Diagnosis we found out that only Visit is a significant predictor if we are concern about how MLU changes over time.

### [OPTIONAL] Exercise 5) Comment on how the three linguistic variables measure linguistic performance (the so-called “construct validity” of the measures). Do they express the same variance?