Anaphora Project Update

Jasmine Falk

July 20, 2020

1 Brief Review

Hypotheses

1. Children improve in anaphora resolution ability as they age
2. Children exhibit consistent mechanisms to resolve anaphora
3. Parent speech reflects child language abilities

Follow-up from 6/14

1. Because the split anaphora data was processed incorrectly (each object was counted as a separate anaphoric instance instead of being counted as multiple objects being referred to by a single anaphor), the data must be processed again (see Section 2).
2. All metrics (resolution accuracy scores, total number of anaphora, total number of split anaphora, percentage of one/split/pronominal anaphora) must be recalculated and graphs plotted again (see Section 3).

2 Error in Counting Split Anaphora

Wrote script to fix the issue with the split anaphora. Data that was previously recorded like this:

	A	B	C	D	E	F	G	H
1	subID	onset	offset	refID	cue	type	prop-target	prop-other
44	1202	292.8	293.83	8	1	2	0	0.8755
45	1202	292.8	293.83	17	1	2	0	0.8755
46	1202	292.8	293.83	24	1	2	0	0.8755
47	1202	299.95	302.14	8	1	2	0	0.5359
48	1202	299.95	302.14	17	1	2	0	0.5359
49	1202	299.95	302.14	24	1	2	0.5359	0
50	1202	308.62	309.65	8	1	2	0	1
51	1202	308.62	309.65	17	1	2	0	1
52	1202	308.62	309.65	24	1	2	1	0
53	1202	325.22	326.11	8	1	2	0	0.65374
54	1202	325.22	326.11	17	1	2	0	0.65374
55	1202	325.22	326.11	24	1	2	0	0.65374

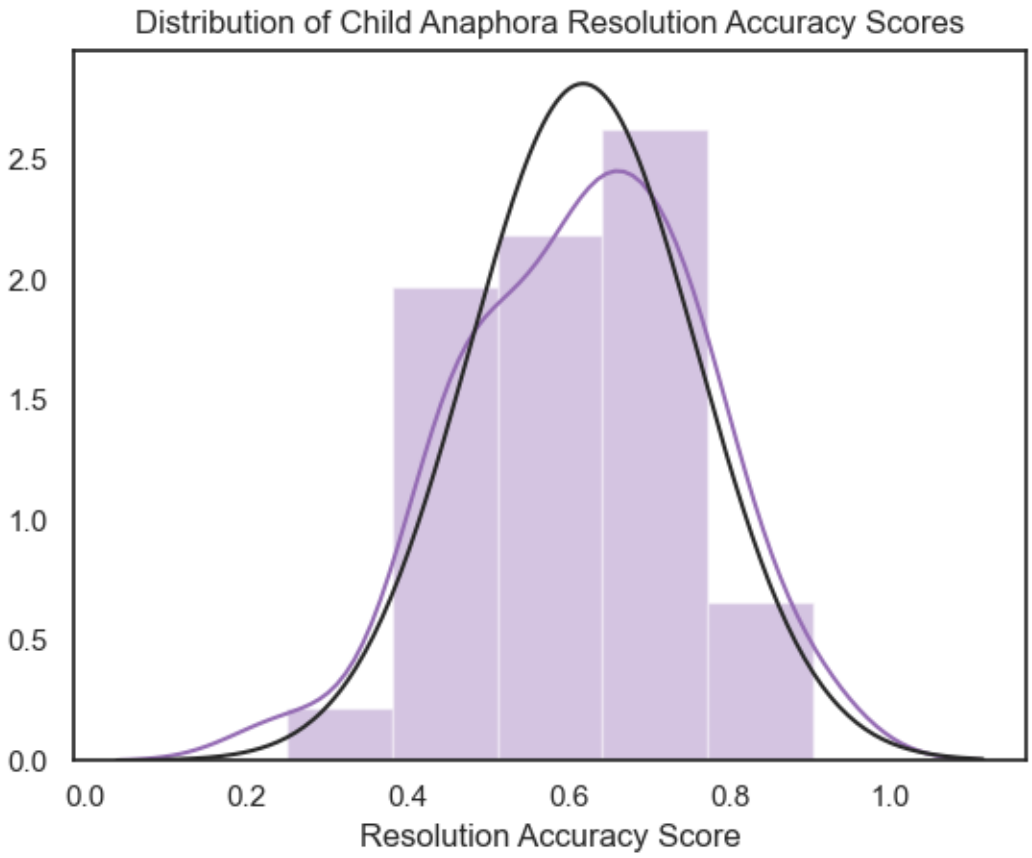
now looks like this:

	A	B	C	D	E	F	G	H
1	subID	onset	offset	refID	cue	type	prop-target	prop-other
44	1202	292.8	293.83	8, 17, 24	1	2	0	1
45	1202	299.95	302.14	8, 17, 24	1	2	0.5359	0.4641
46	1202	308.62	309.65	8, 17, 24	1	2	1	0
47	1202	325.22	326.11	8, 17, 24	1	2	0	1

(**notes**: color boxes added in images to delineate separate split anaphora. `prop-target` indicates the proportion of time between `onset` and `offset` that the subject spent fixating on the target object, `prop-other` indicates the proportion of time the subject spent fixating on an object other than the target).

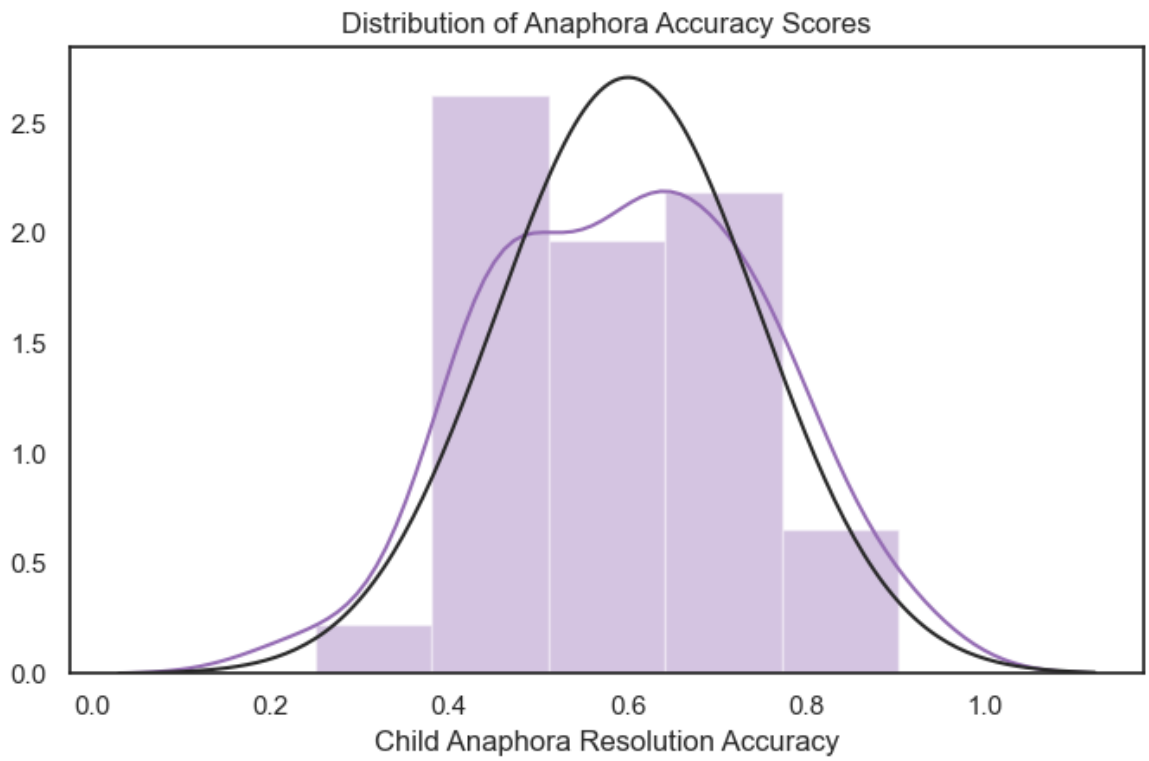
3 New Results

3.1 Distribution of resolution accuracy scores



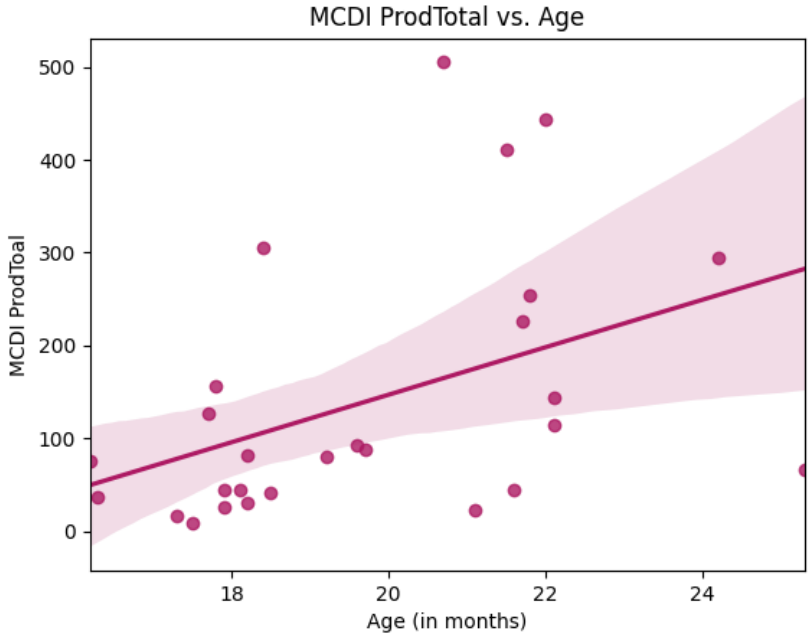
MEAN	0.61783566
STDEV	0.14362333
MEDIAN	0.63
MAX	0.90322581
MIN	0.25

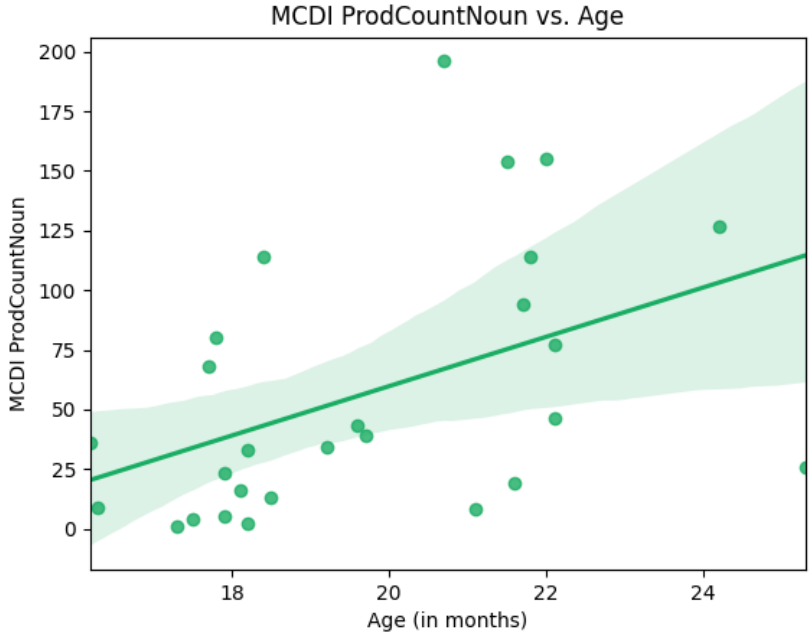
For reference, prior to fixing the split anaphora error, the scores were overall lower. This is expected given that the error was falsely lowering scores, especially for subjects who used a lot of split anaphora.



MEAN	0.60047840
STDEV	0.14941899
MEDIAN	0.6095238
MAX	0.90322581
MIN	0.25

3.2 Child age vs. MCDI scores





3.3 Percentage of total utterances containing anaphora

MEAN	28.75362336
STDEV	11.2244521
MEDIAN	29.8
MAX	53.54330709
MIN	0.0

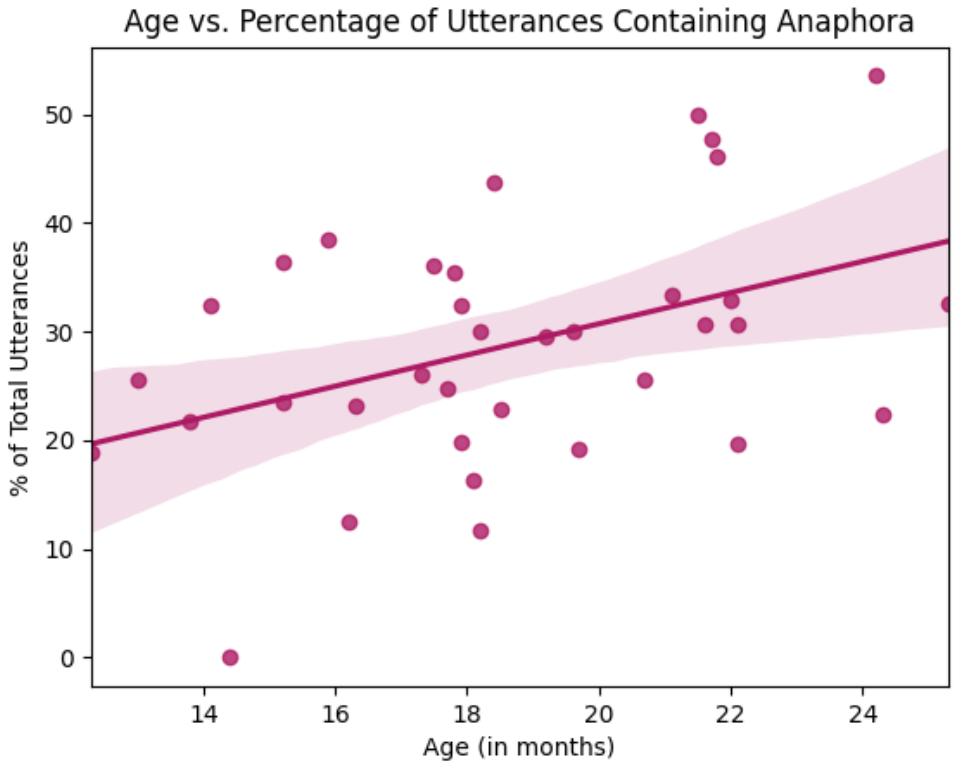


Figure 1: Compared to child age

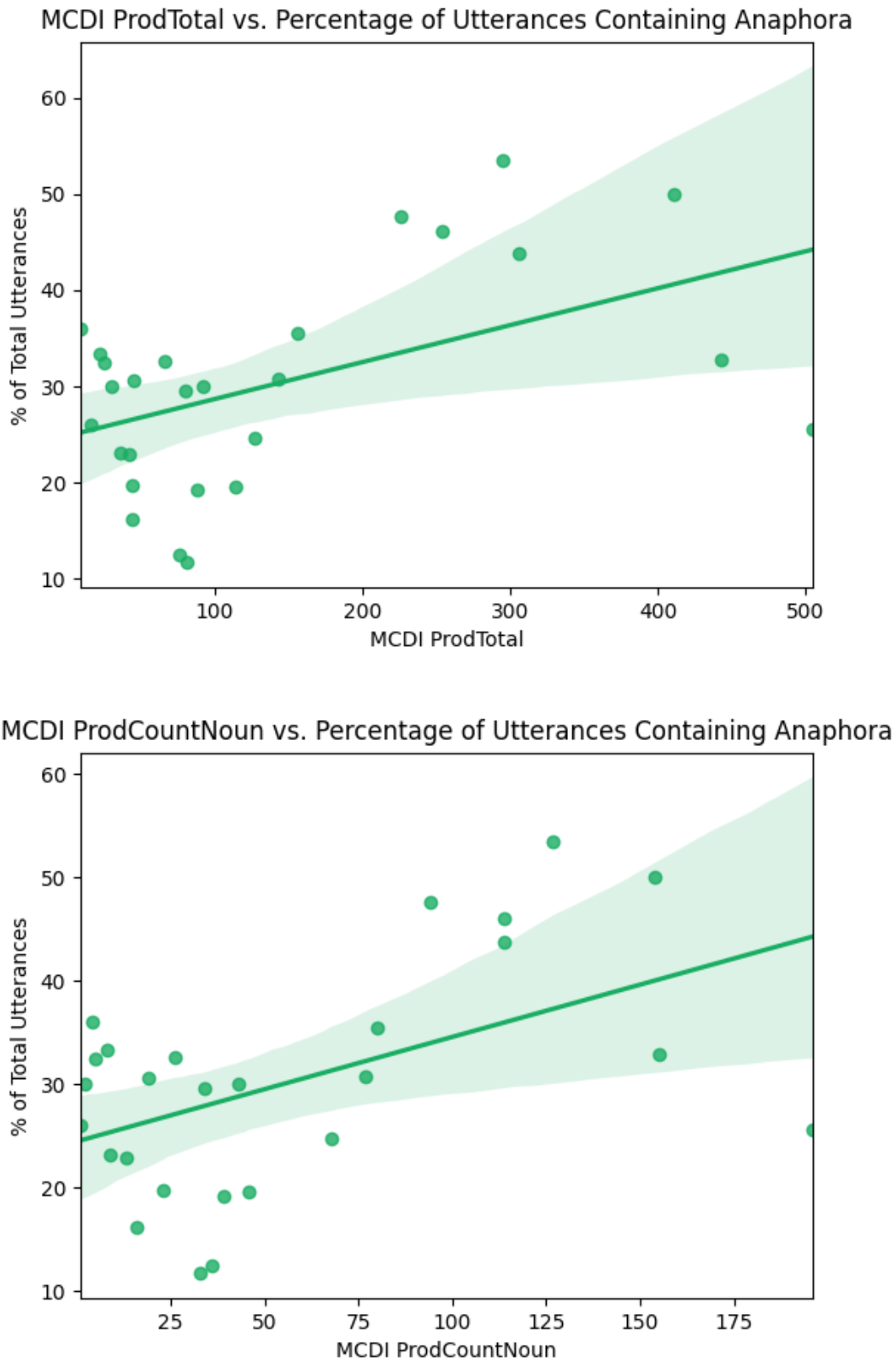


Figure 2: Compared to MCDI scores

3.4 Percentage of anaphoric utterances containing split anaphora

MEAN	3.818098971
STDEV	6.975501347
MEDIAN	0.0
MAX	27.1186441
MIN	0.0

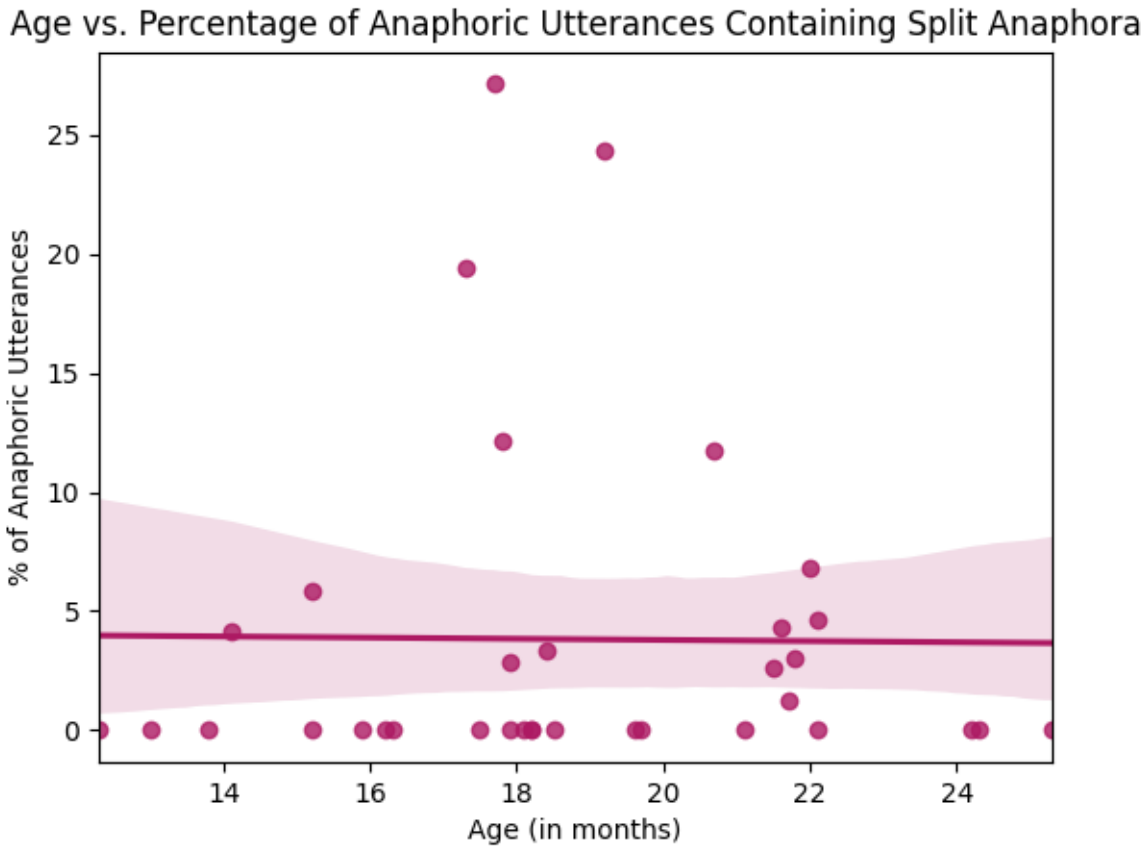
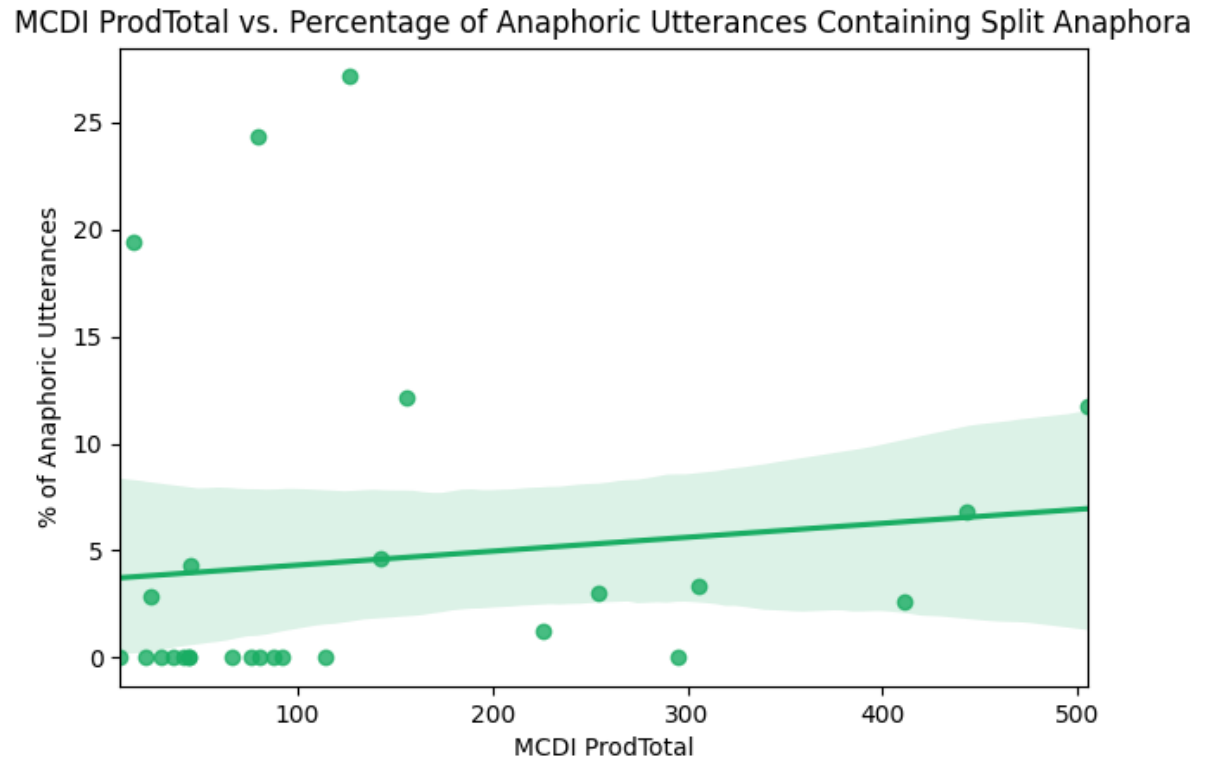


Figure 3: Compared to child age



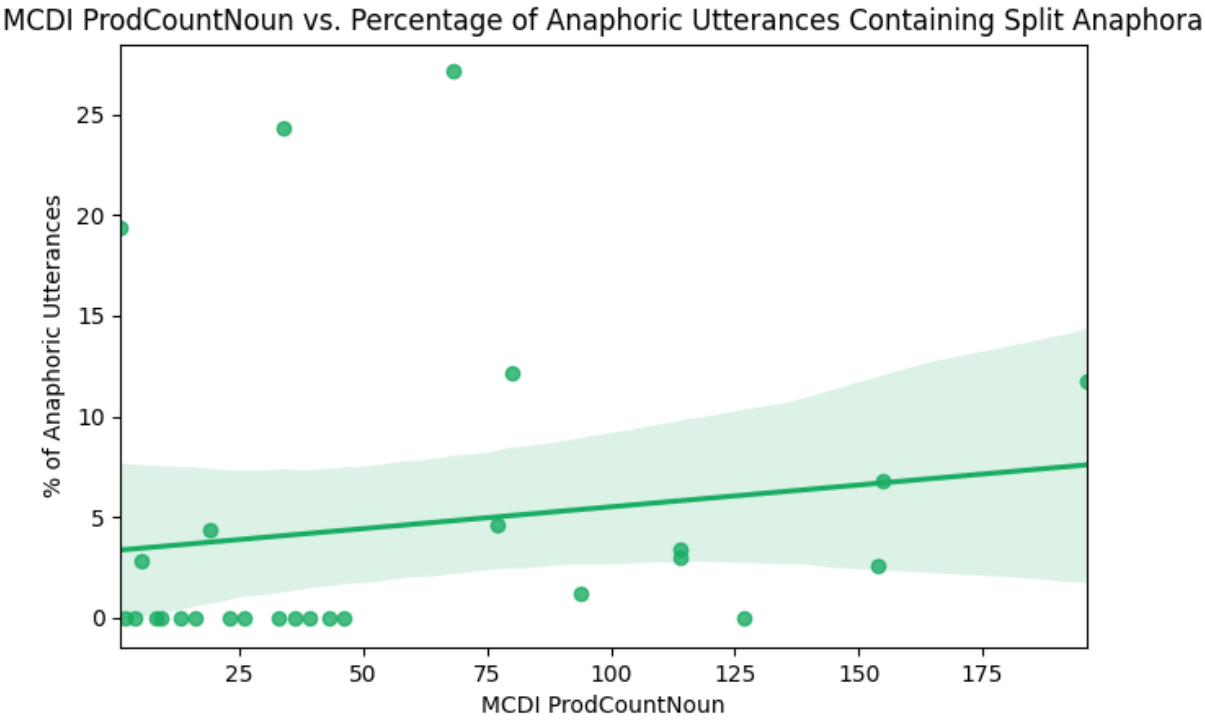


Figure 4: Compared to MCDI scores

3.5 Percentage of anaphoric utterances containing one anaphora

MEAN	4.992586177
STDEV	5.16138845
MEDIAN	4.3478261
MAX	18.1818182
MIN	0.0

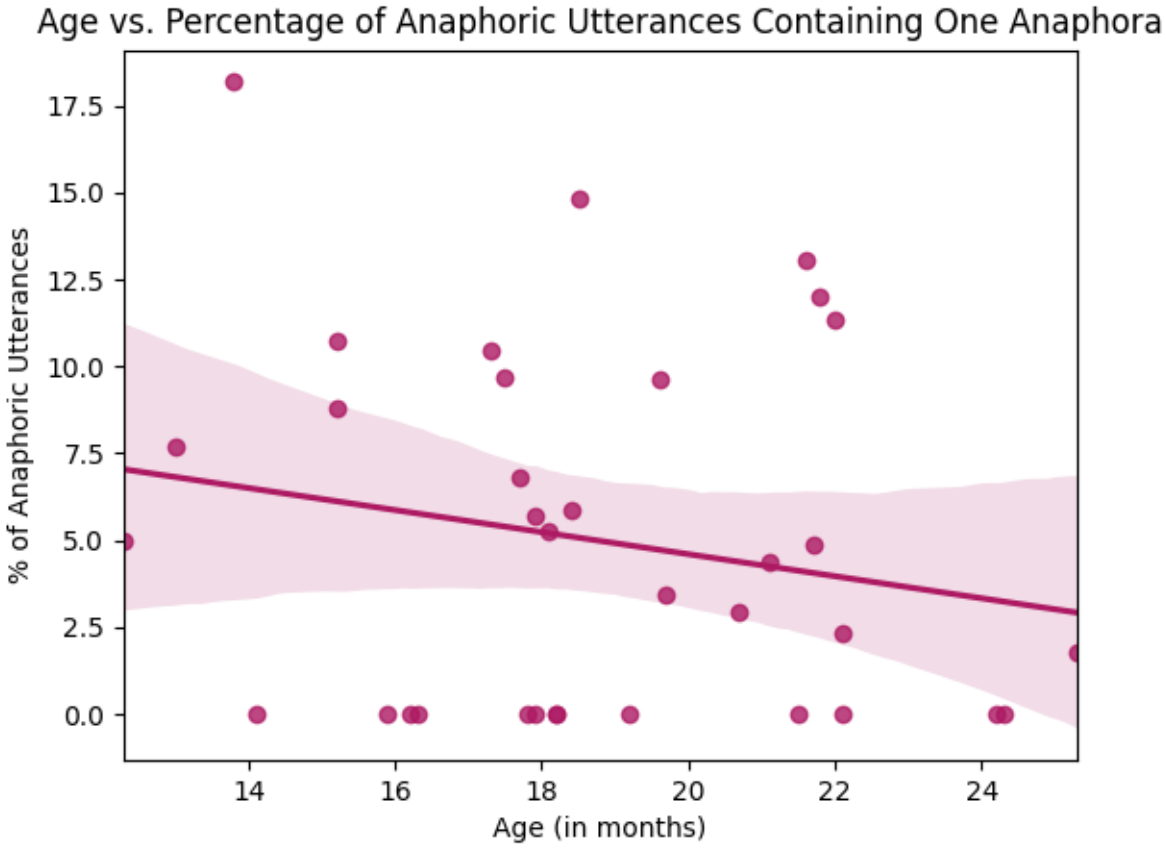


Figure 5: Compared to child age

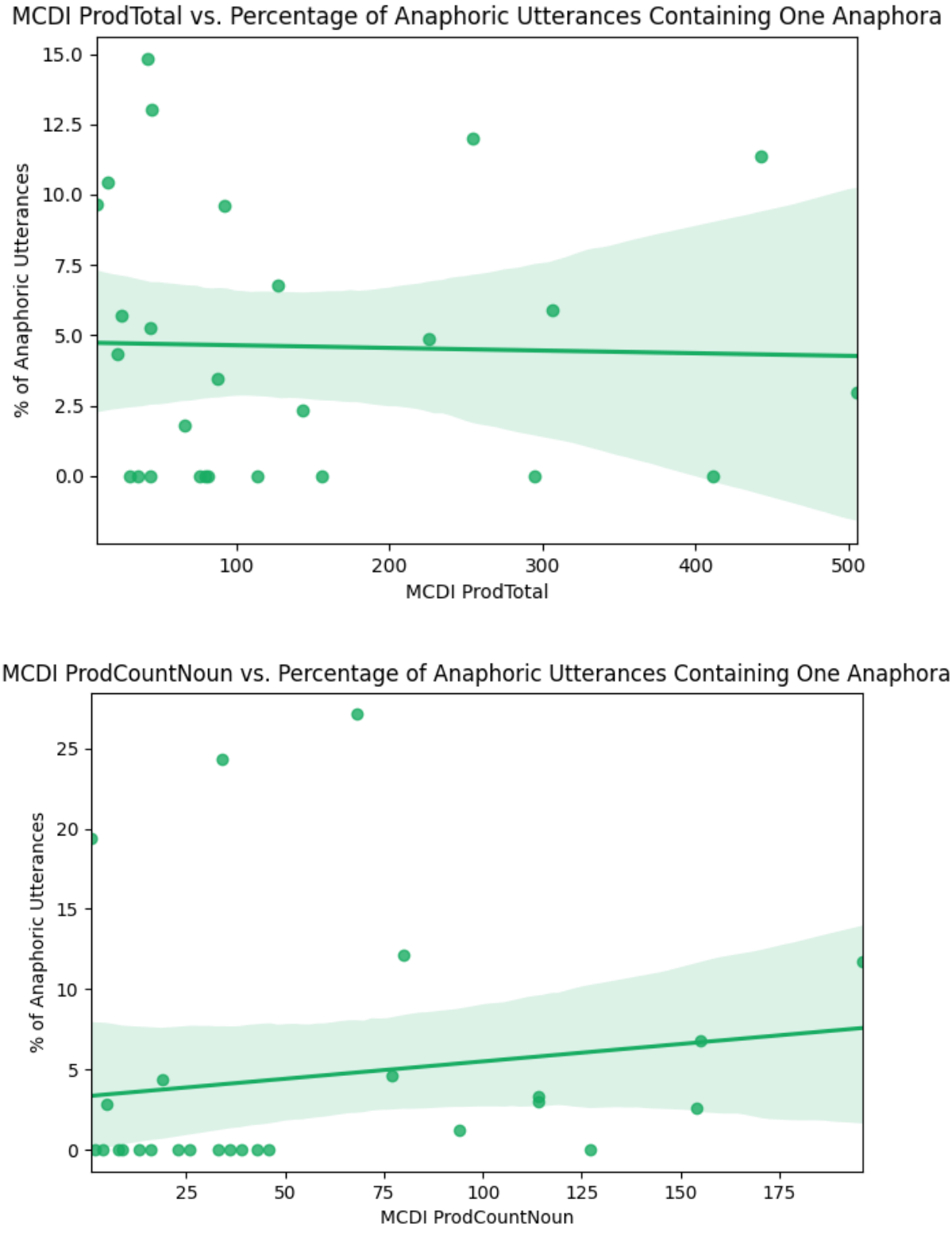


Figure 6: Compared to MCDI scores

3.6 Percentage of anaphoric utterances containing verbally-cued anaphora

A positive correlation between age and the percentage of utterances containing verbally-cued anaphora can be found. This is expected given that parents would likely assume an older child would be able to comprehend anaphora with fewer non-verbal cues (e.g. pointing). A positive correlation, although smaller, can still be observed in the MCDI score graphs. This corroborates the findings in Section 3.3, which found that there is a positive correlation between age and percentage of utterances containing anaphora, potentially suggesting that parents perceive older children as more competent language users, specifically with regards to anaphora resolution.

MEAN	0.619767878
STDEV	0.174723248
MEDIAN	0.625
MAX	1.0
MIN	0.260869565

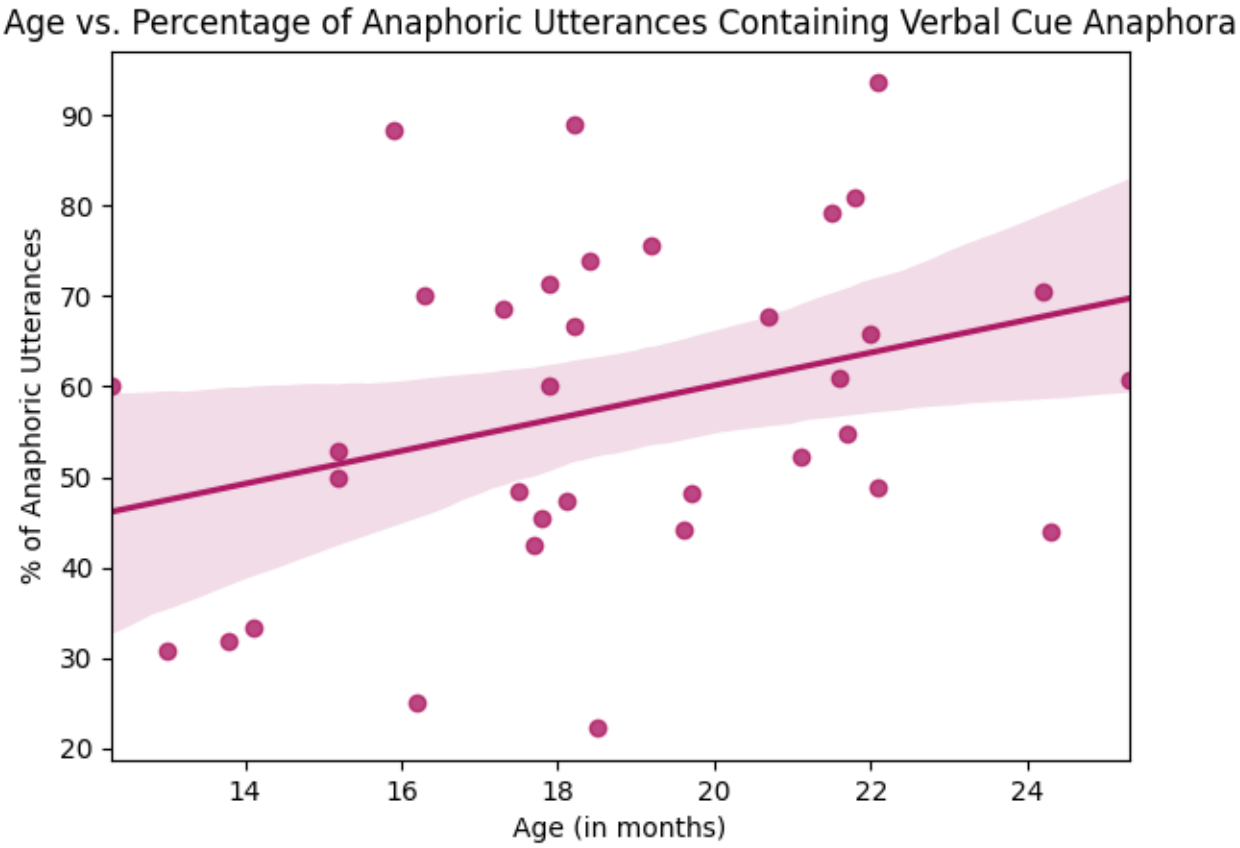


Figure 7: Compared to child age

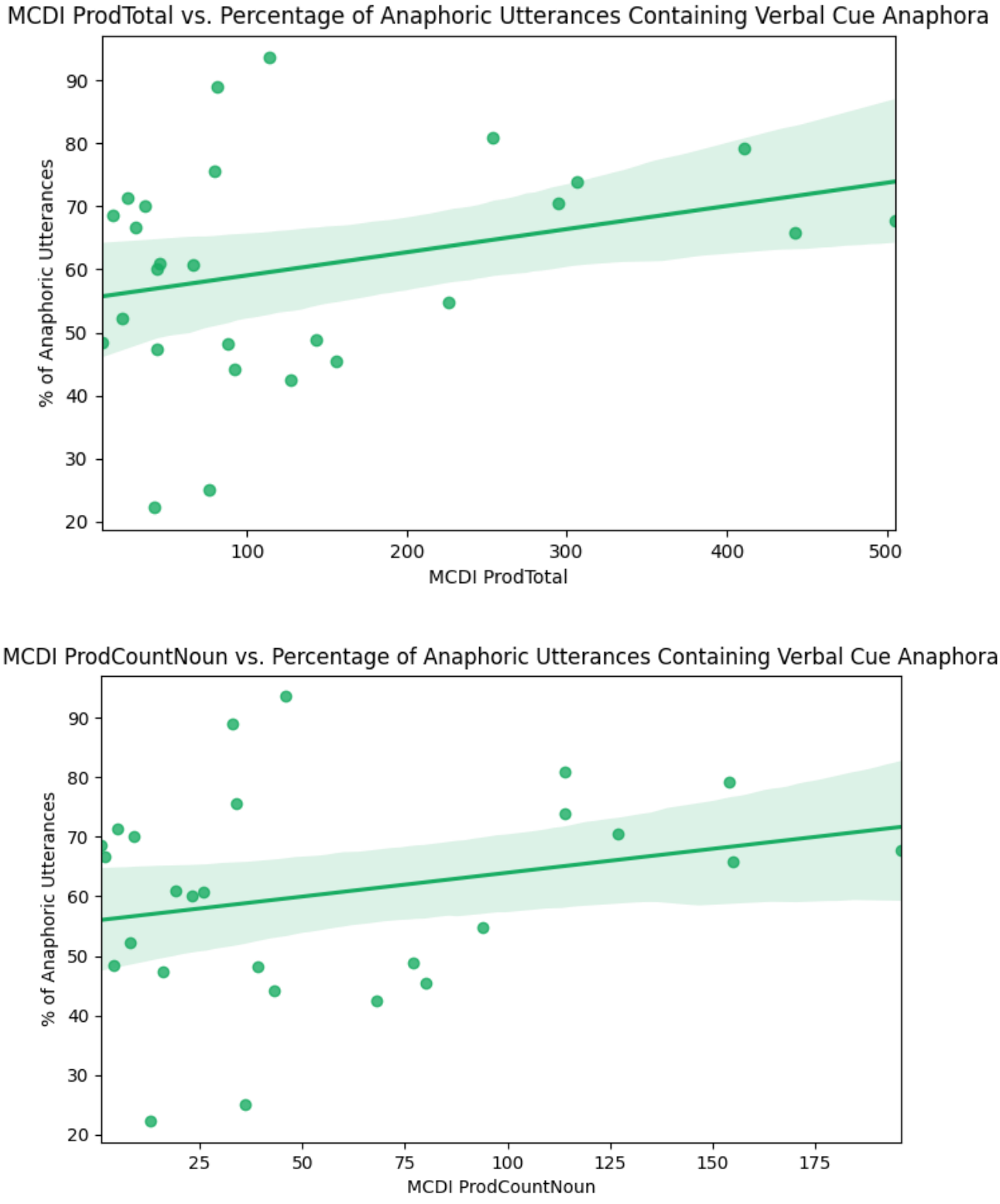


Figure 8: Compared to MCDI scores

3.7 Percentage of anaphoric utterances containing visually-cued anaphora

A negative correlation between age and the percentage of utterances containing visually-cued anaphora can be found. This is expected given that parents would likely assume a younger child would require more extra-textual cues to resolve anaphora (e.g. pointing). A negative correlation, although smaller, can still be observed in the MCDI score graphs. These findings are expected given the results of Section 3.6.

MEAN	0.613708164
STDEV	0.18033212
MEDIAN	0.636363636
MAX	1.0
MIN	0.166666667

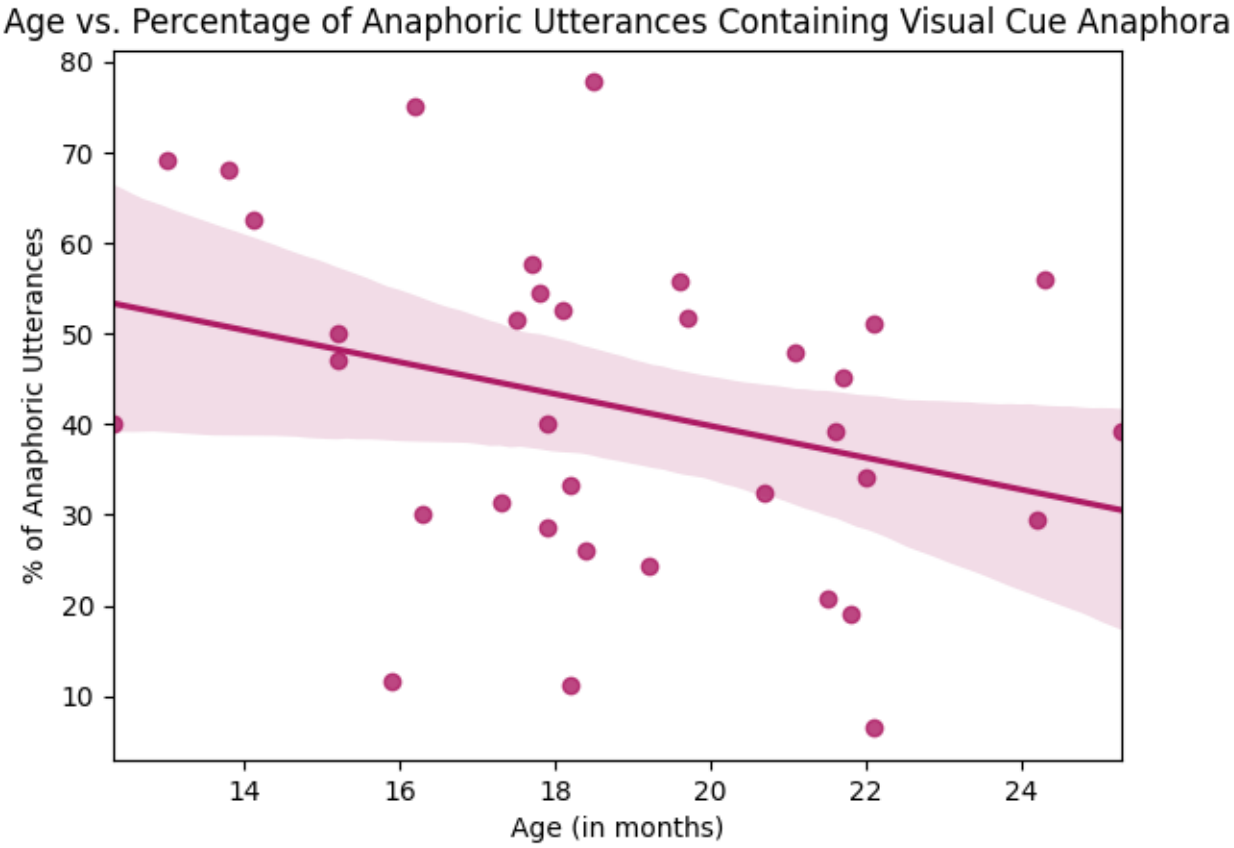


Figure 9: Compared to child age

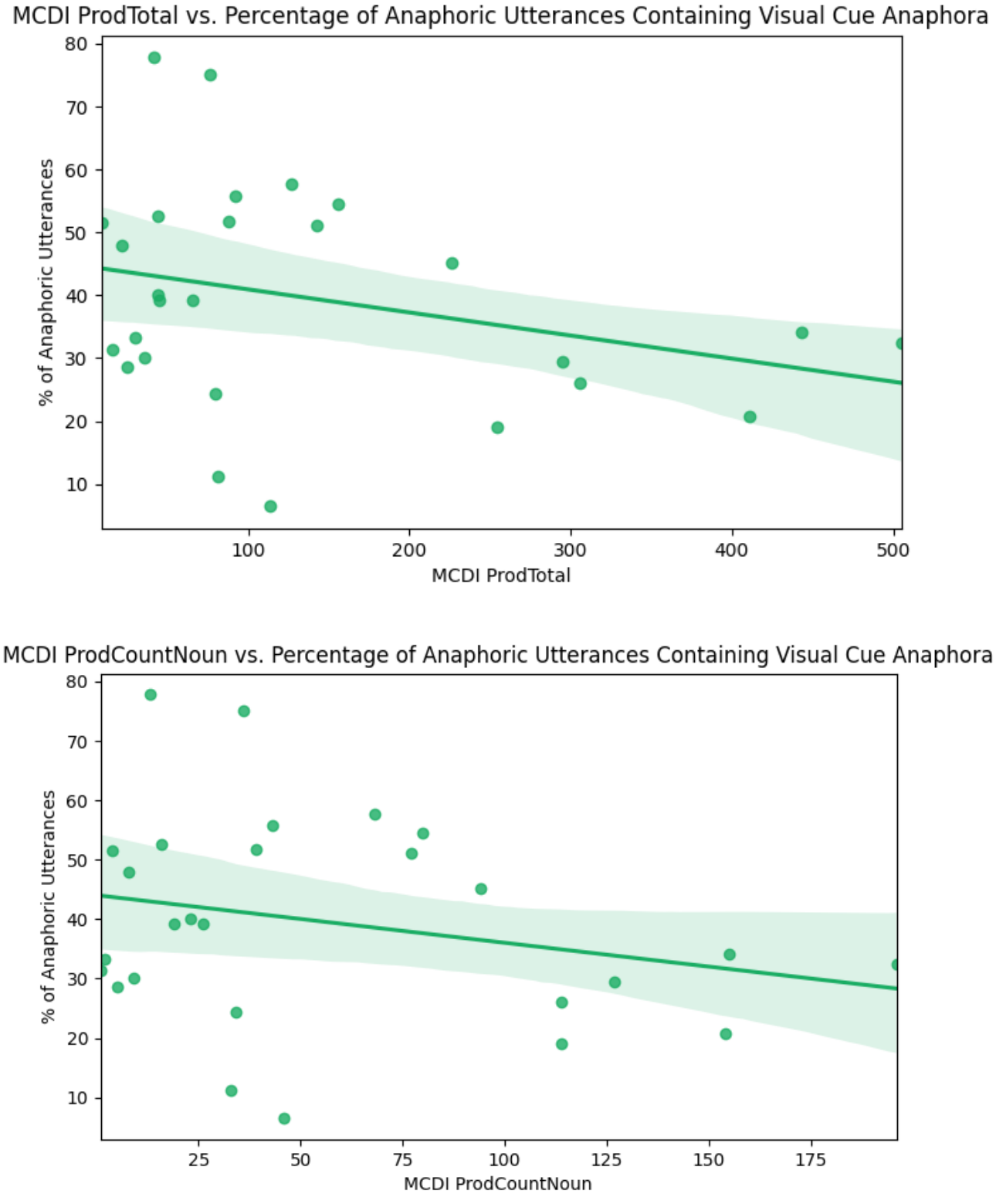


Figure 10: Compared to MCDI scores

3.8 Anaphora resolution accuracy score

MEAN	0.61783566
STDEV	0.14362333
MEDIAN	0.63
MAX	0.90322581
MIN	0.25

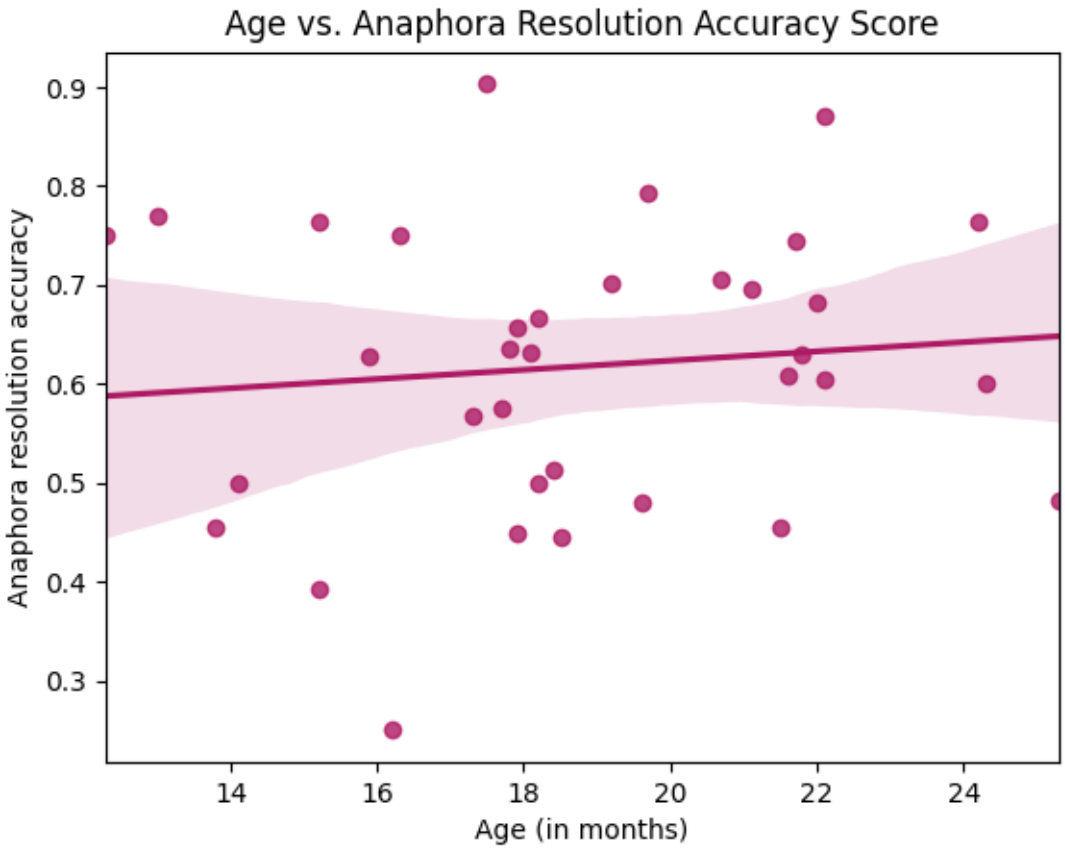
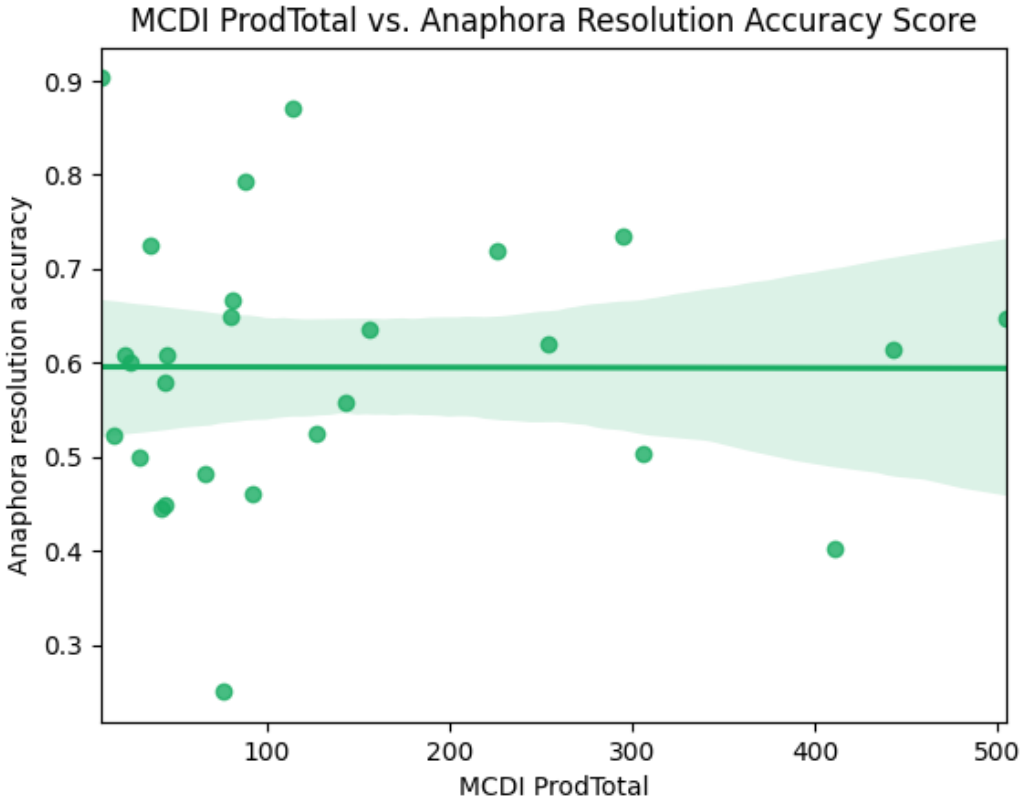


Figure 11: Compared to child age



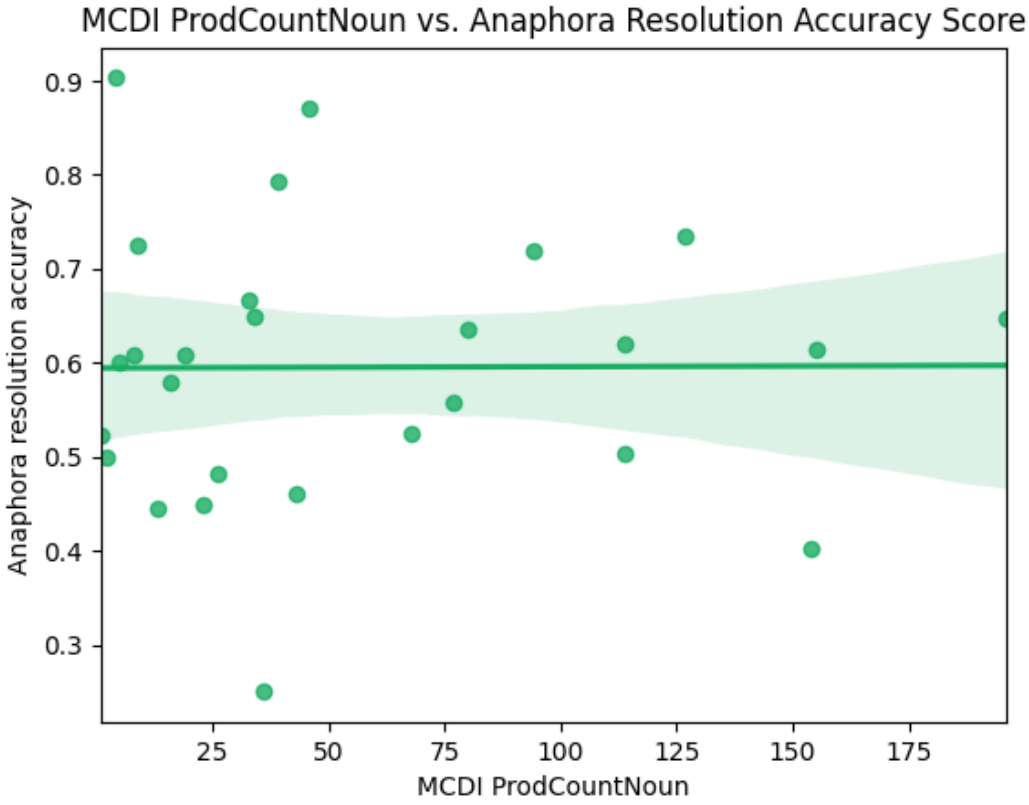


Figure 12: Compared to MCDI scores

3.9 Pronominal anaphora resolution accuracy score

MEAN	0.623075598
STDEV	0.156496398
MEDIAN	0.611111111
MAX	0.892857143
MIN	0.25

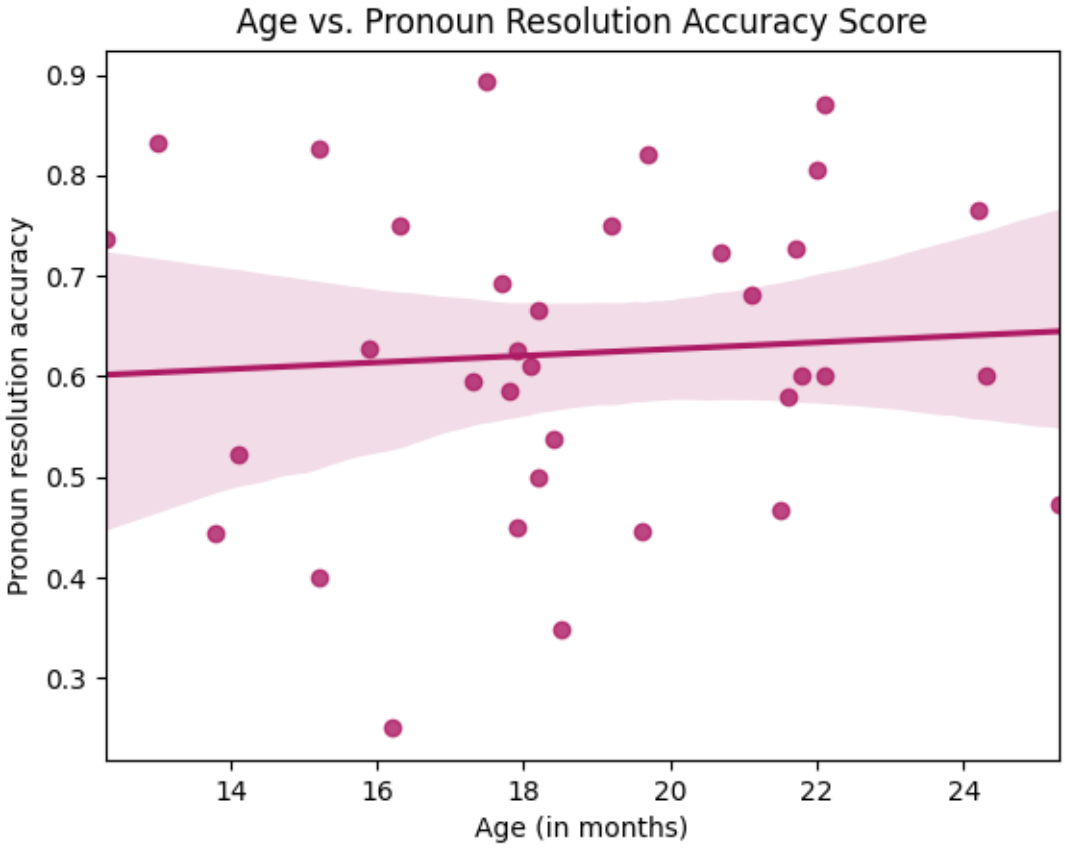
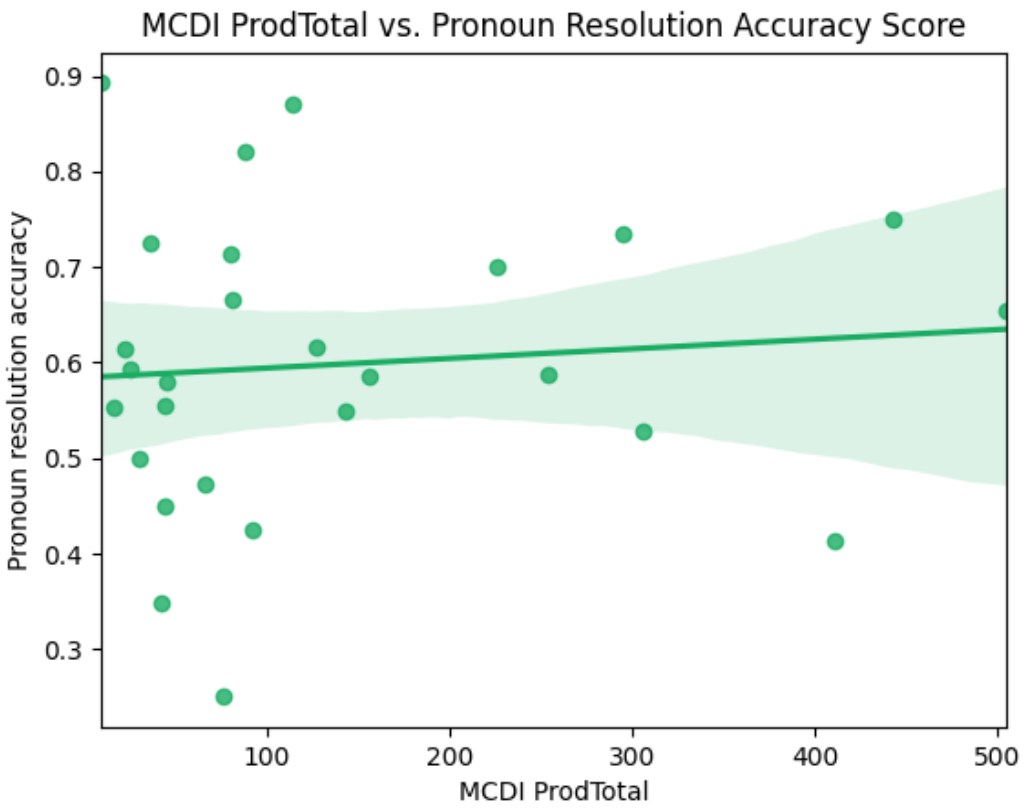


Figure 13: Compared to child age



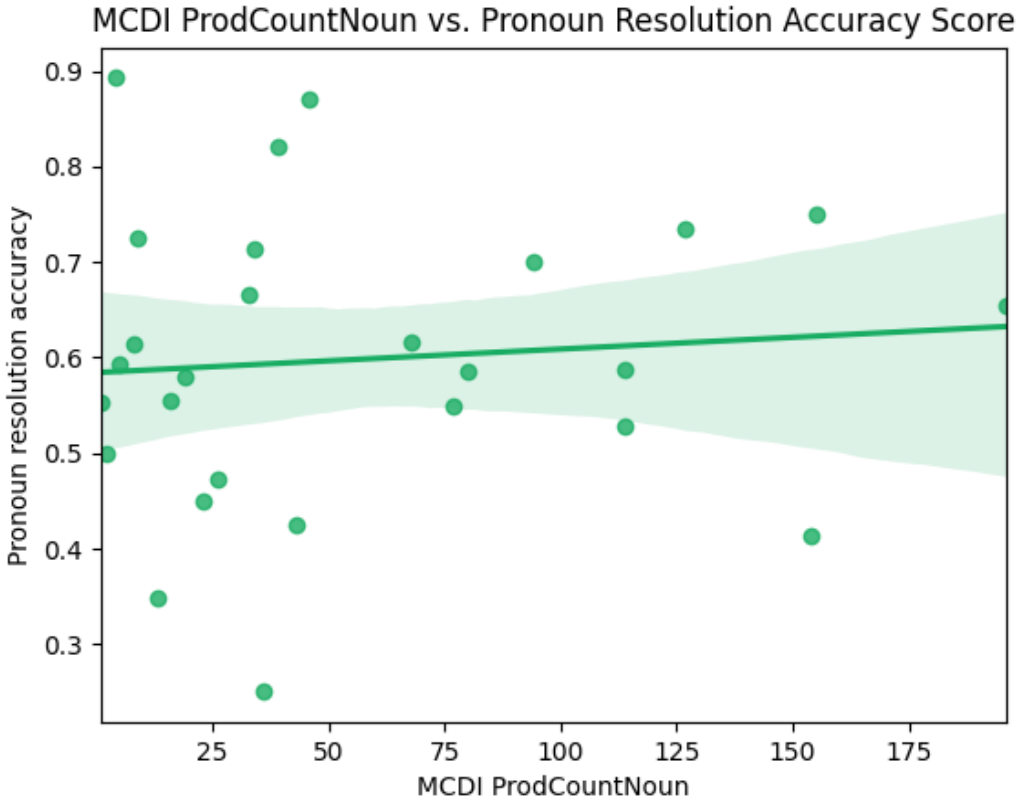


Figure 14: Compared to MCDI scores

3.10 One anaphora resolution accuracy score

Note that not all subjects used one anaphora in their speech, so this data is much more limited.

MEAN	0.707251082
STDEV	0.35626686
MEDIAN	0.845238095
MAX	1.0
MIN	0.0

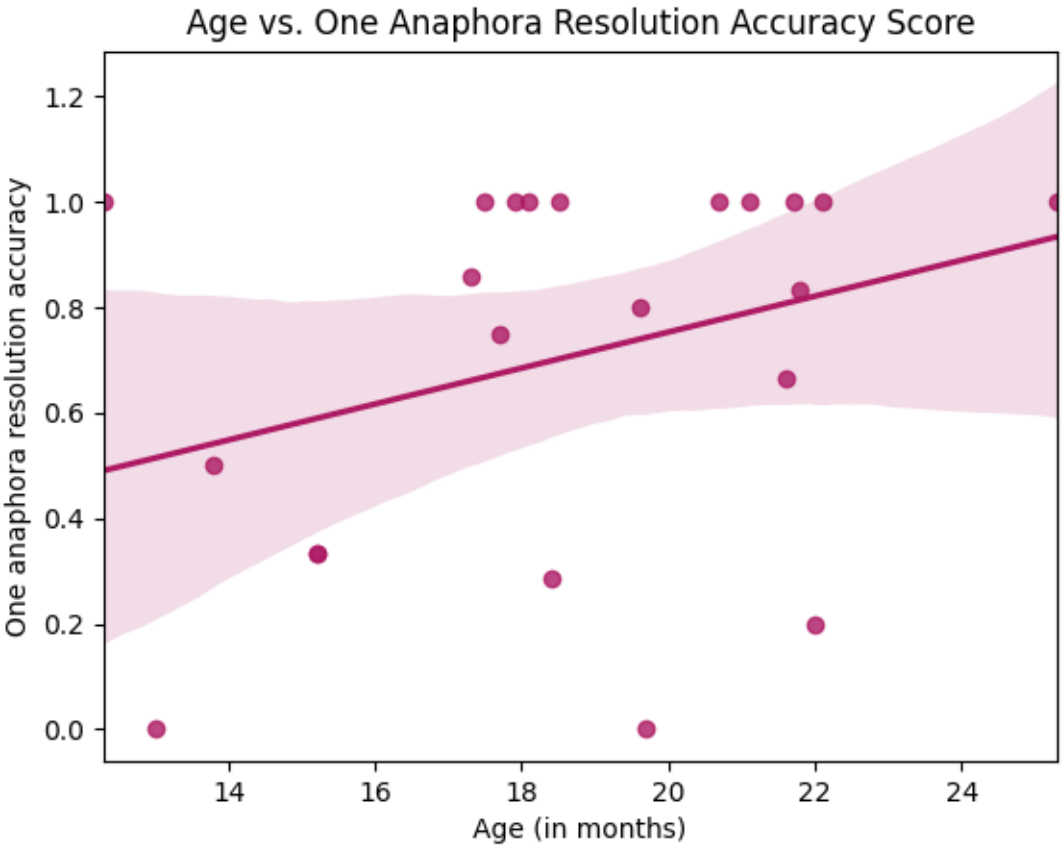
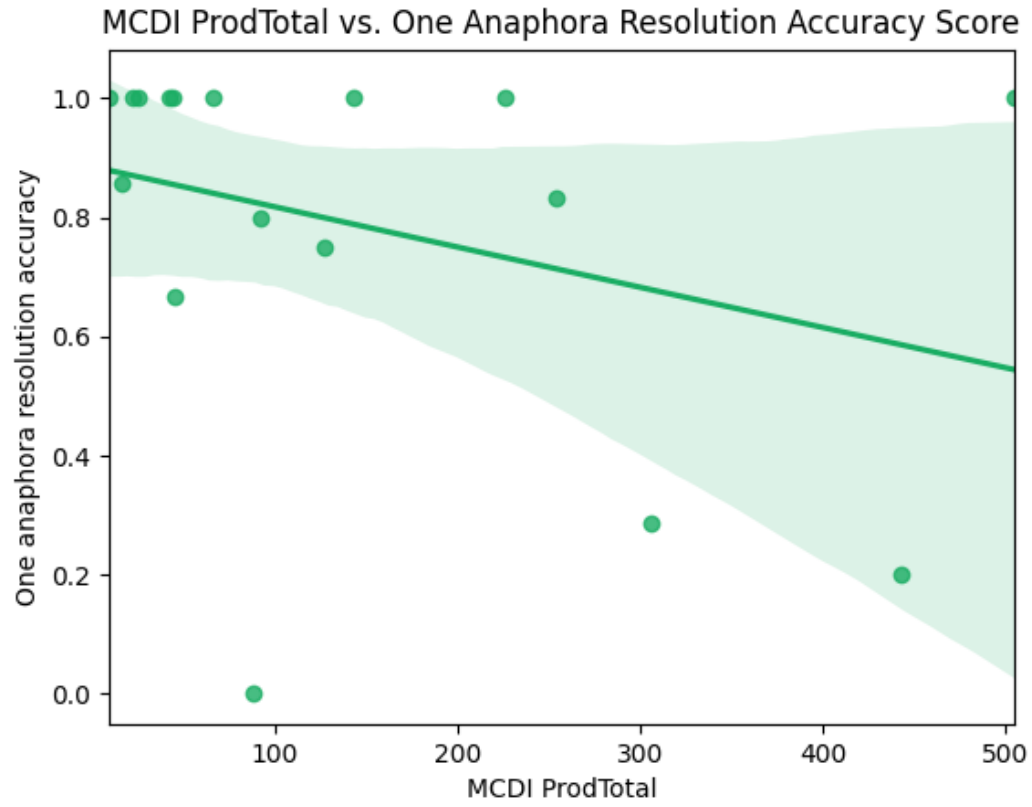


Figure 15: Compared to child age



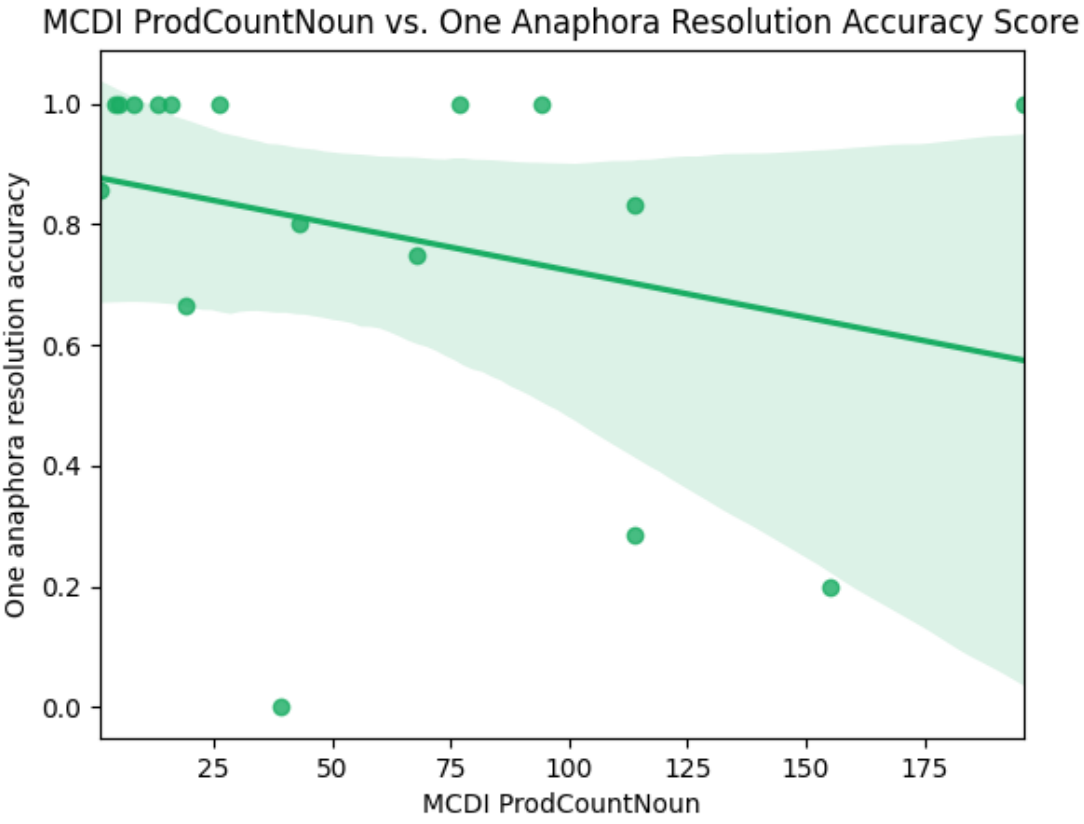


Figure 16: Compared to MCDI scores

3.11 Split anaphora resolution accuracy score

Note that not all subjects used split anaphora in their speech, so this data is much more limited.

MEAN	0.501994302
STDEV	0.372519107
MEDIAN	0.5
MAX	1.0
MIN	0.0

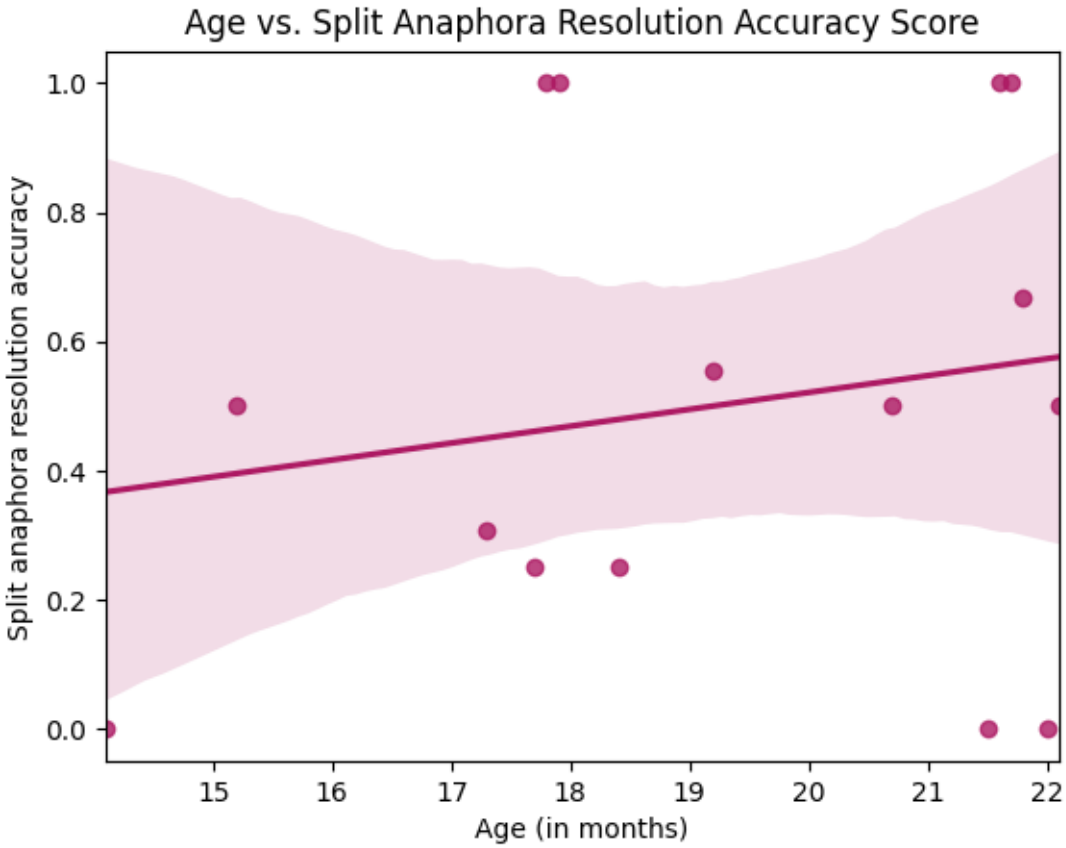


Figure 17: Compared to child age

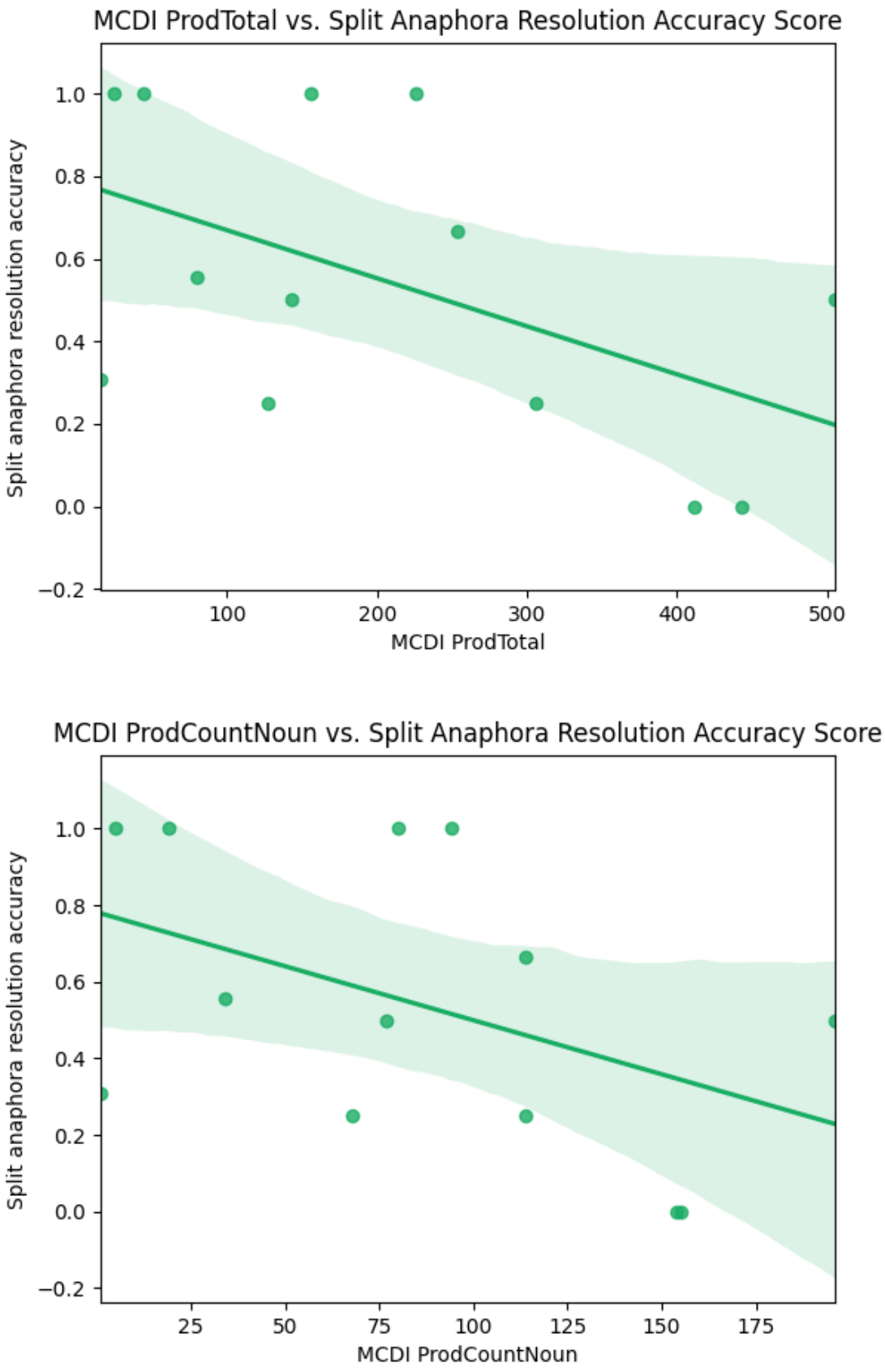


Figure 18: Compared to MCDI scores

3.12 Verbally-cued anaphora resolution accuracy score

MEAN	0.619767878
STDEV	0.174723248
MEDIAN	0.625
MAX	1.0
MIN	0.260869565

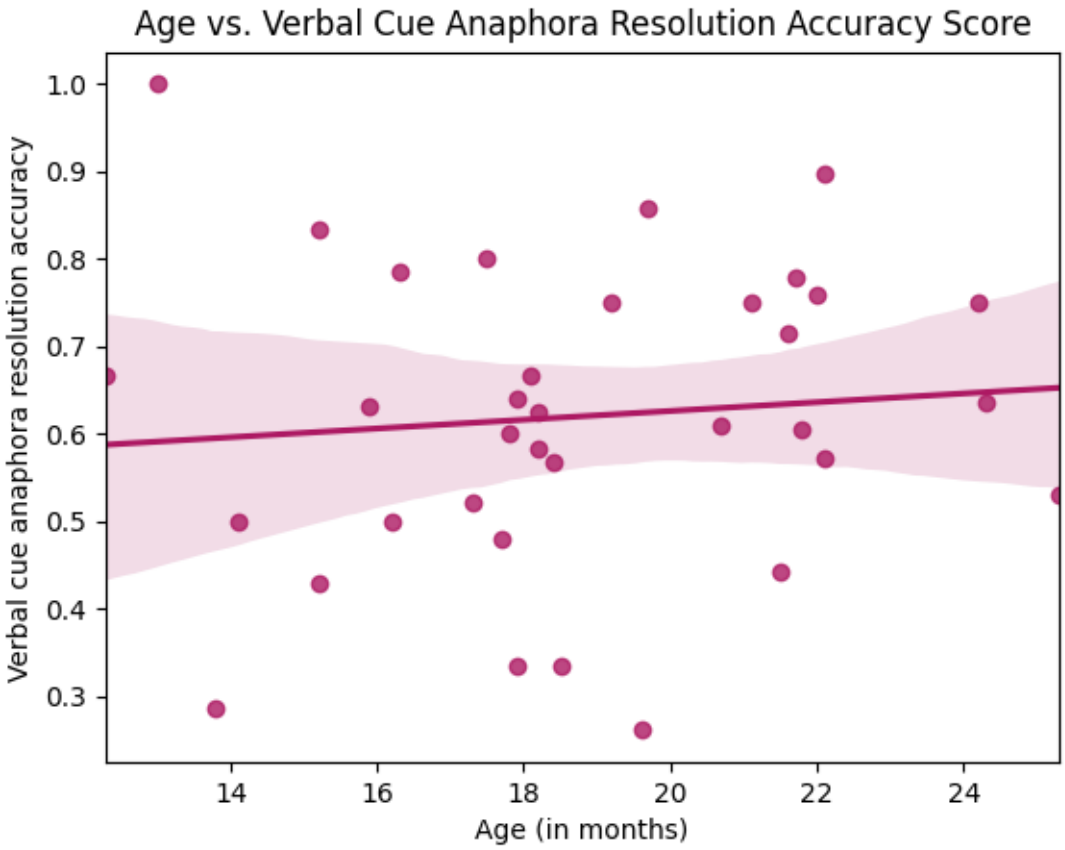
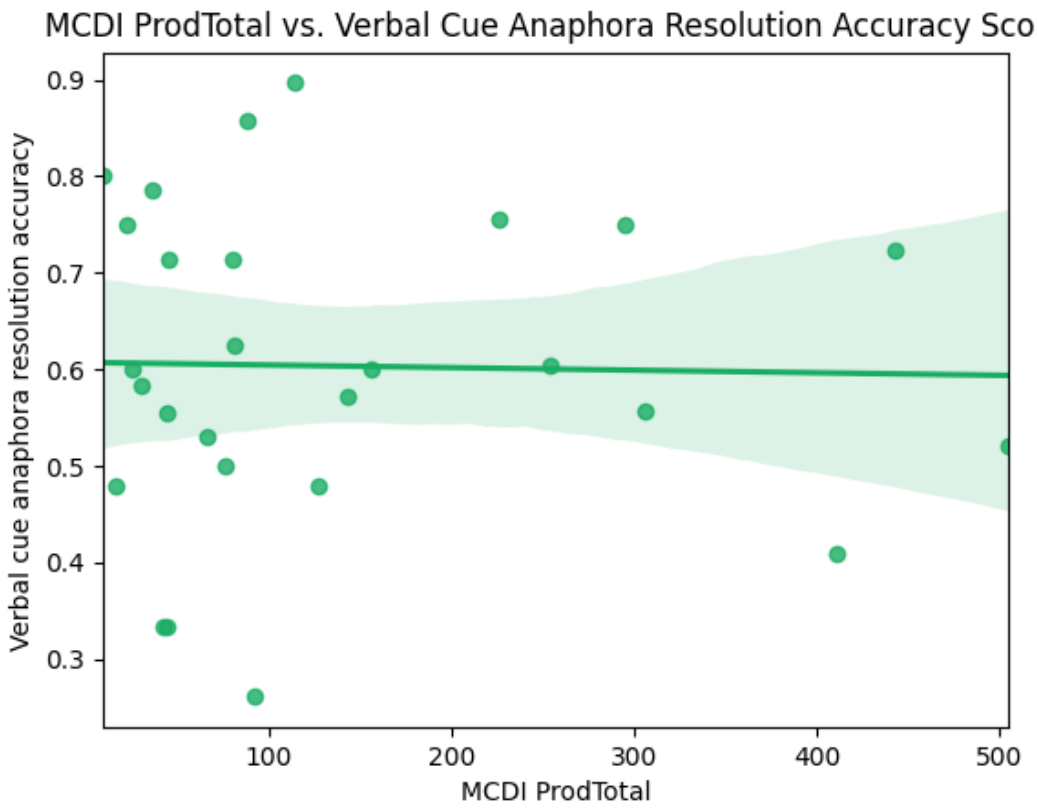


Figure 19: Compared to child age



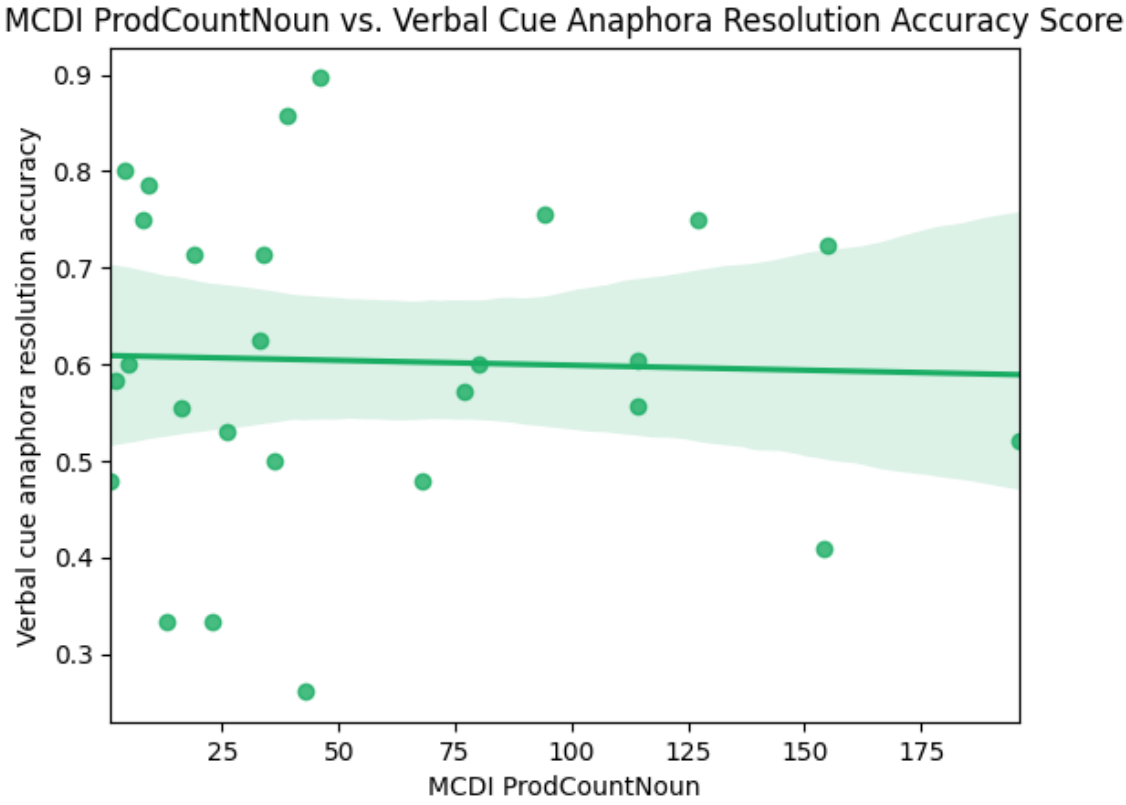


Figure 20: Compared to MCDI scores

3.13 Visually-cued anaphora resolution accuracy score

MEAN	0.613708164
STDEV	0.18033212
MEDIAN	0.636363636
MAX	1.0
MIN	0.166666667

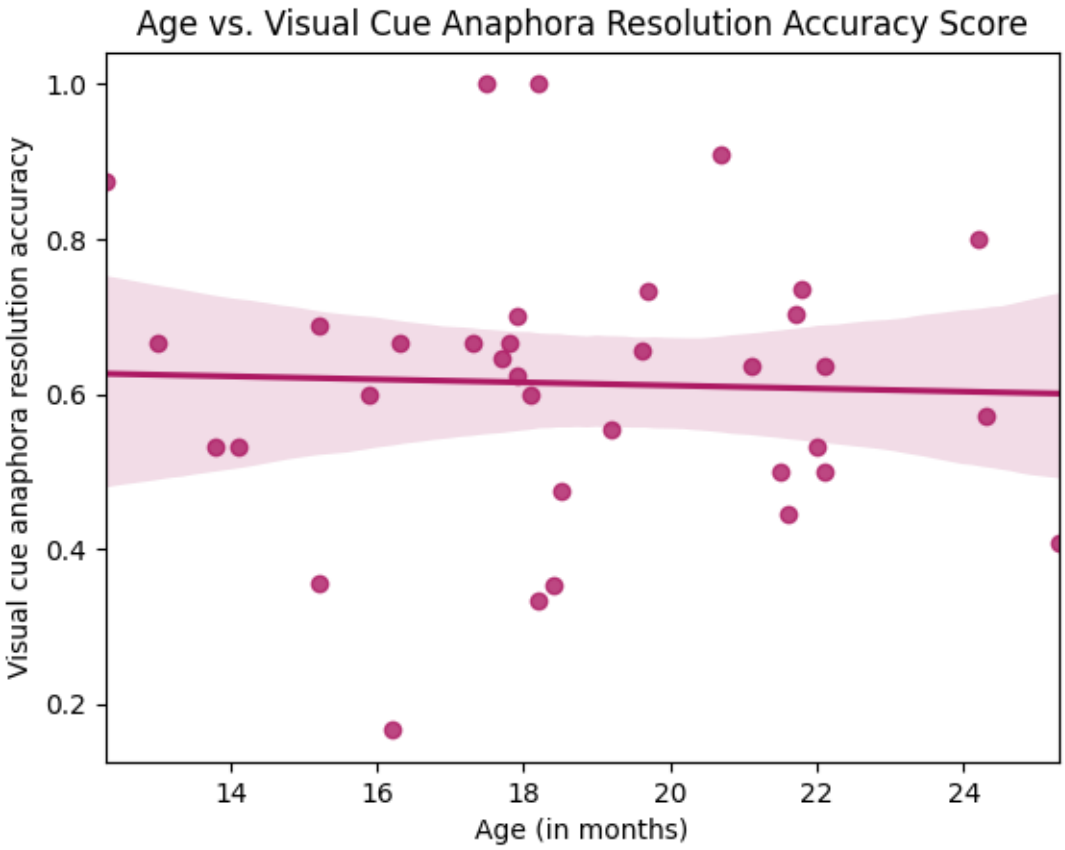
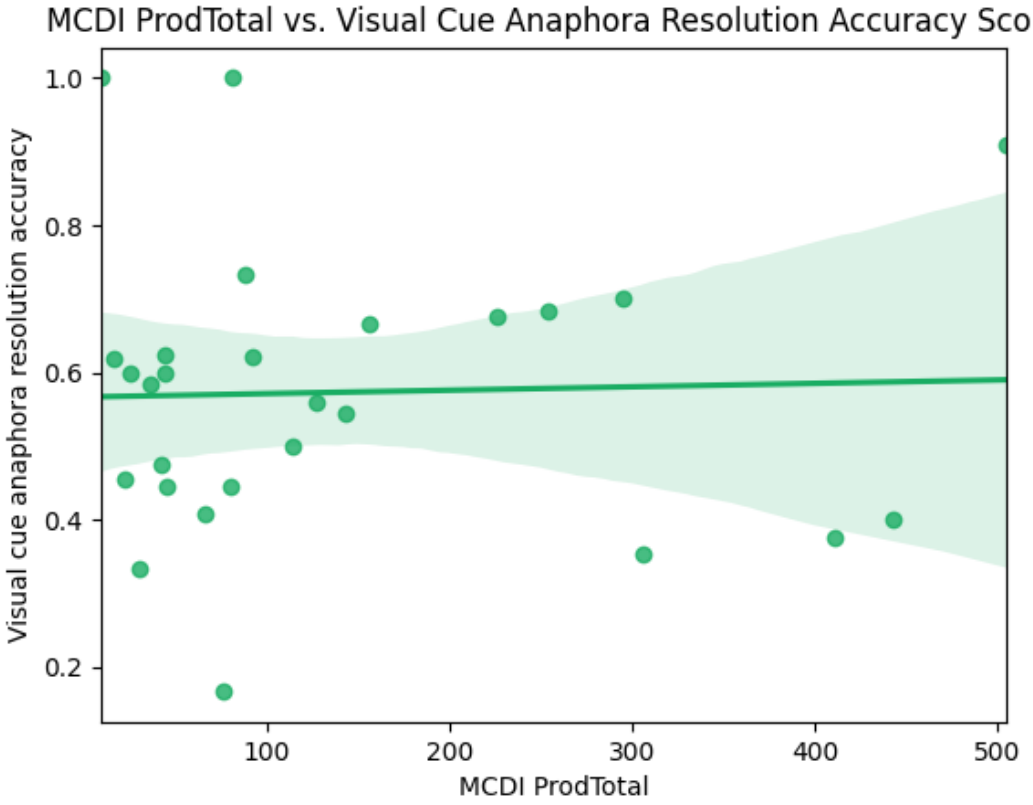


Figure 21: Compared to child age



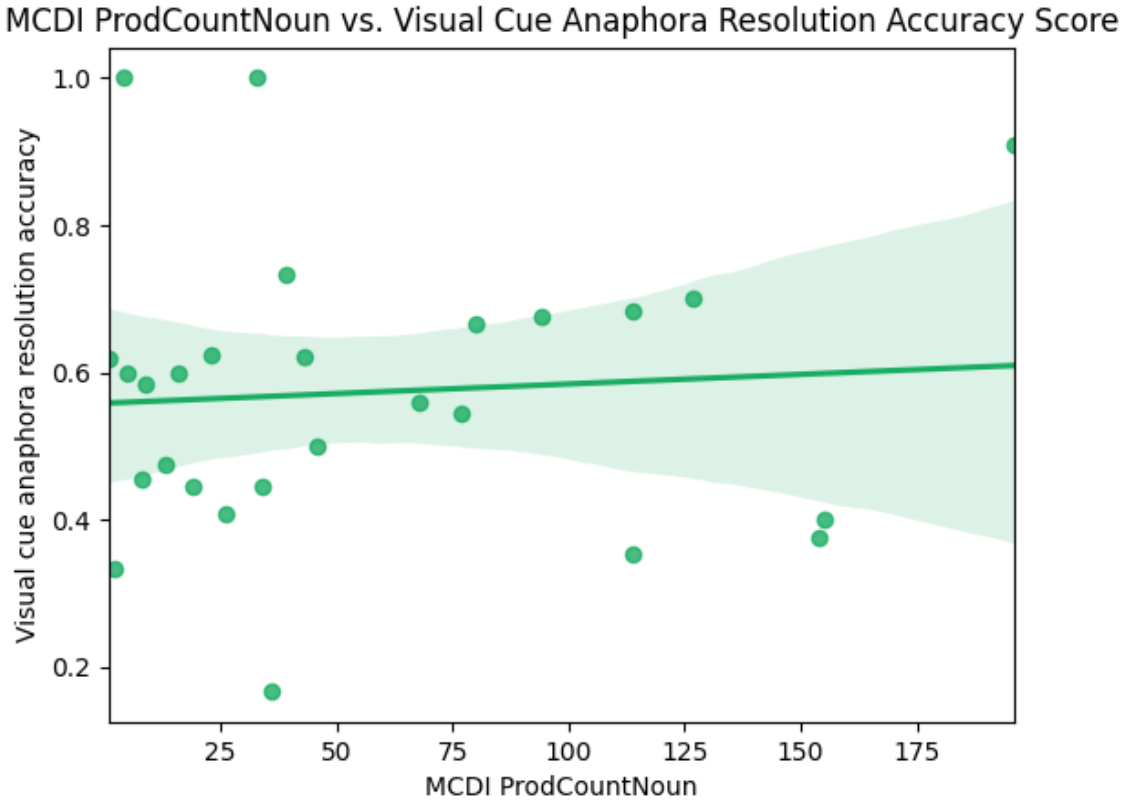


Figure 22: Compared to MCDI scores

4 Summary

Split Anaphora Error

Data was reprocessed to account for error in how the split anaphora data was represented. The new distribution of anaphora resolution accuracy scores reflects this fix as expected (see Section 3.1), showing that the new scores were overall higher (original mean=0.6005, new mean=0.6178).

New Plots

The new results, which use the revised resolution accuracy scores and also include MCDI scores and anaphora types as additional covariates, don't seem drastically different from the original results reported in the thesis. Sections 3.3, 3.6, and 3.7 report potentially interesting findings, but beyond that, few expected correlations were found, even when compared to MCDI scores.

Next Steps

1. Eliminate instances of child-driven anaphora and then analyzing the data again.
2. Determine whether parent gaze is a factor in where children attend to should be investigated.
3. Analysis of the storybook dataset for anaphora to determine whether this dataset could be better for this project.