

Anaphora Project Update

Jasmine Falk

July 23, 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 the parent's perception of/assumptions about their child's language comprehension 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

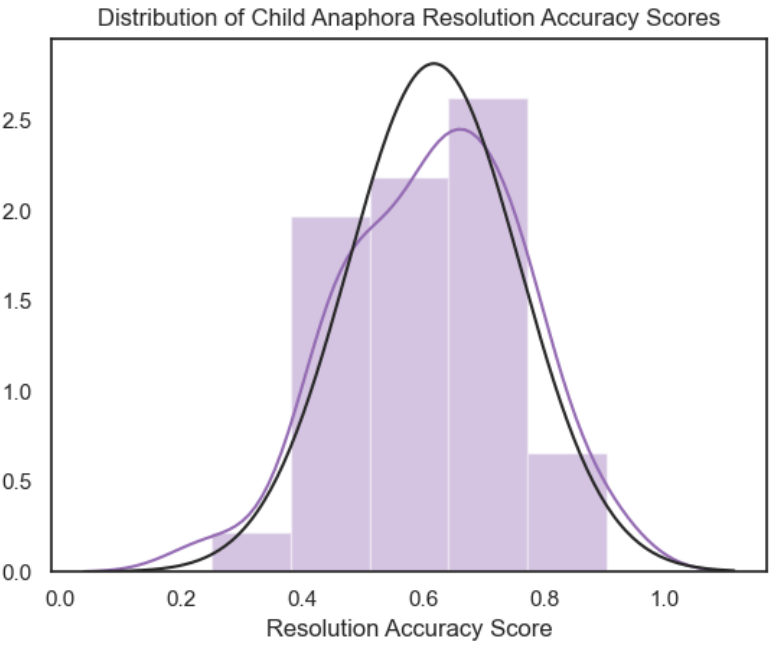
now looks like this:

	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).

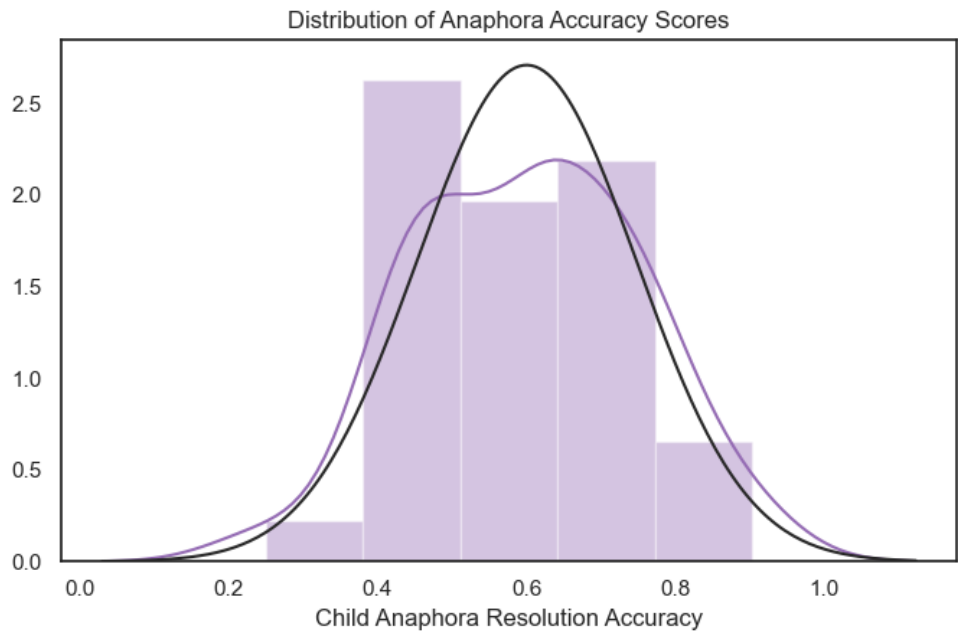
3 New Results

3.1 Distribution of resolution accuracy scores



mean	0.618
stdev	0.144
median	0.630
max	0.903
min	0.250

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	0.600
stdev	0.149
median	0.610
max	0.903
min	0.250

3.2 Child age vs. MCDI scores

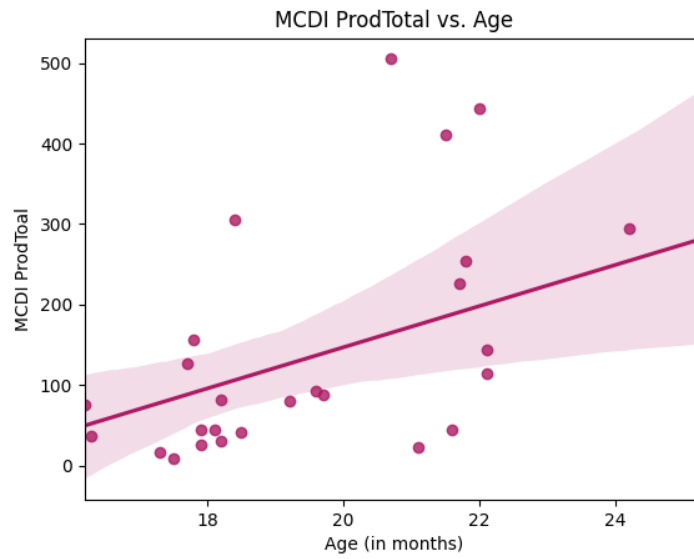


Figure 1: $r(25) = .431, p = .025$.

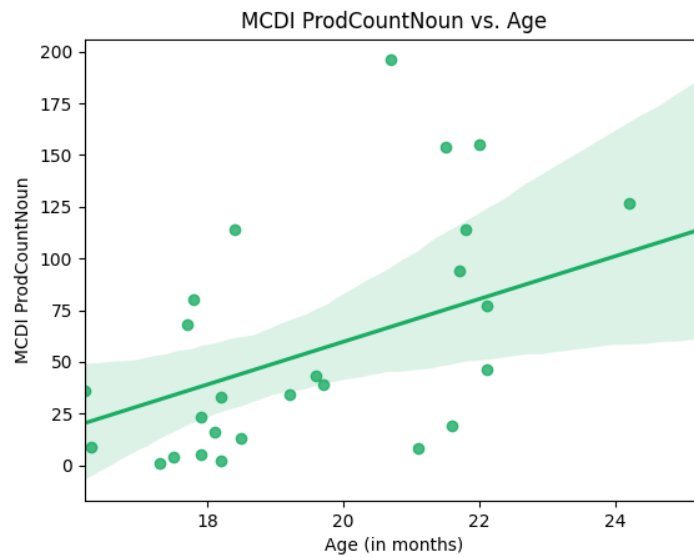


Figure 2: $r(25) = .448, p = .019$.

3.3 Percentage of total utterances containing anaphora

mean	28.75
stdev	11.22
median	29.80
max	53.54
min	0.00

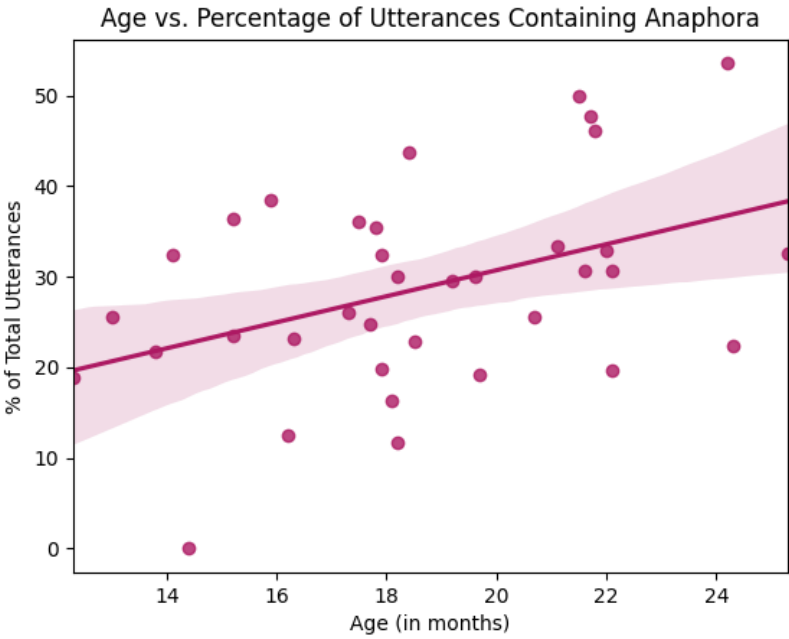


Figure 3: Compared to child age. $r(33) = .367, p = .030$.

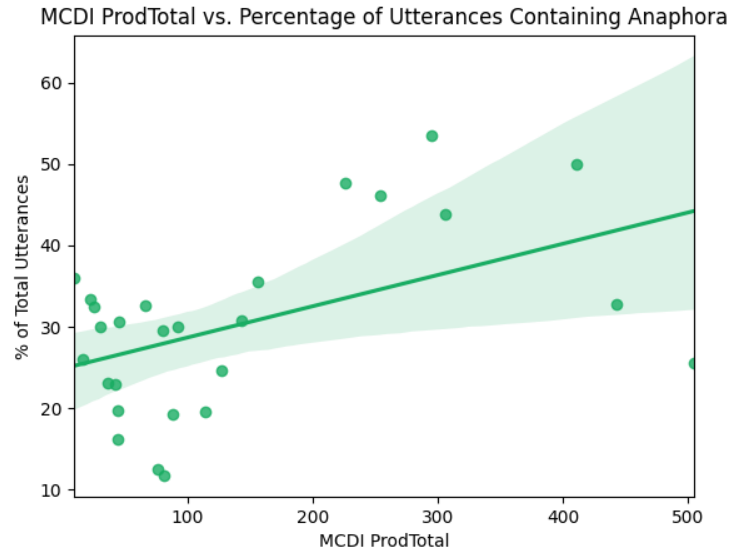


Figure 4: Compared to MCDI ProdTotal scores. $r(25) = .490$, $p = .009$

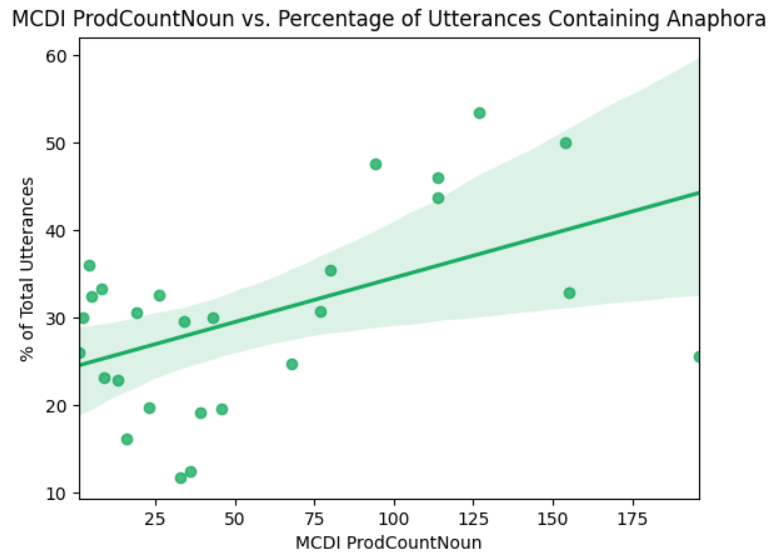


Figure 5: Compared to MCDI ProdCountNoun scores. $r(25) = .504$, $p = .007$

3.4 Percentage of anaphoric utterances containing split anaphora

mean	3.82
stdev	6.98
median	0.00
max	27.12
min	0.00

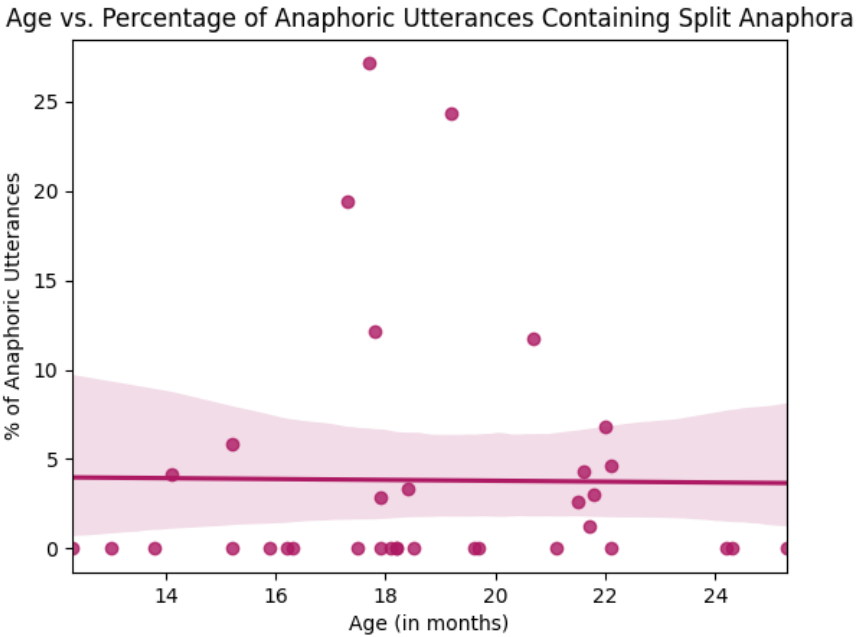


Figure 6: Compared to child age. $r(33) = -.011$, $p = .949$

MCDI ProdTotal vs. Percentage of Anaphoric Utterances Containing Split Anaphora

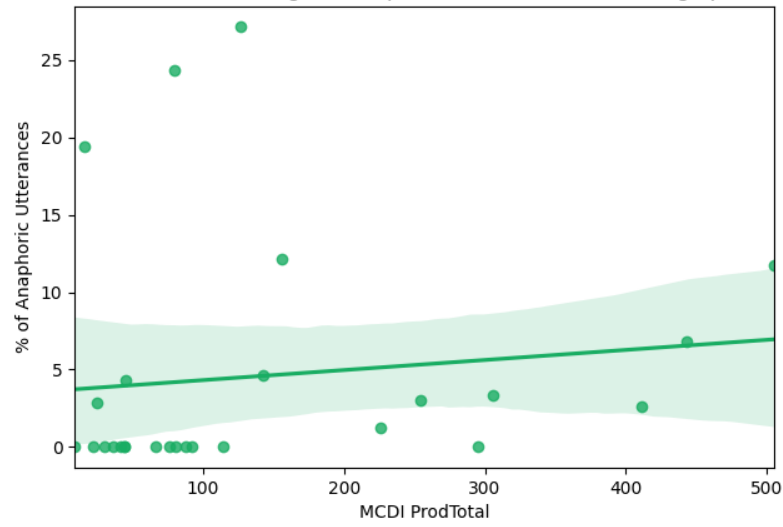


Figure 7: Compared to MCDI ProdTotal scores. $r(25) = .118$, $p = .556$.

MCDI ProdCountNoun vs. Percentage of Anaphoric Utterances Containing Split Anaphora

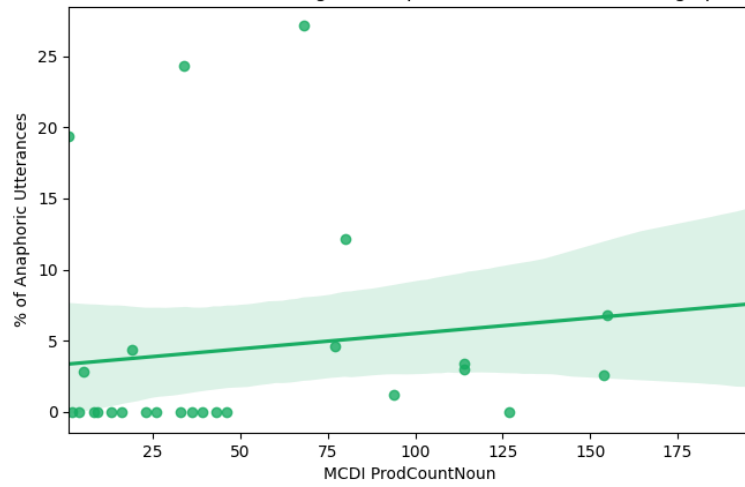


Figure 8: Compared to MCDI ProdCountNoun scores. $r(25) = .154$, $p = .444$.

3.5 Percentage of anaphoric utterances containing one anaphora

mean	4.99
stdev	5.16
median	4.35
max	18.18
min	0.00

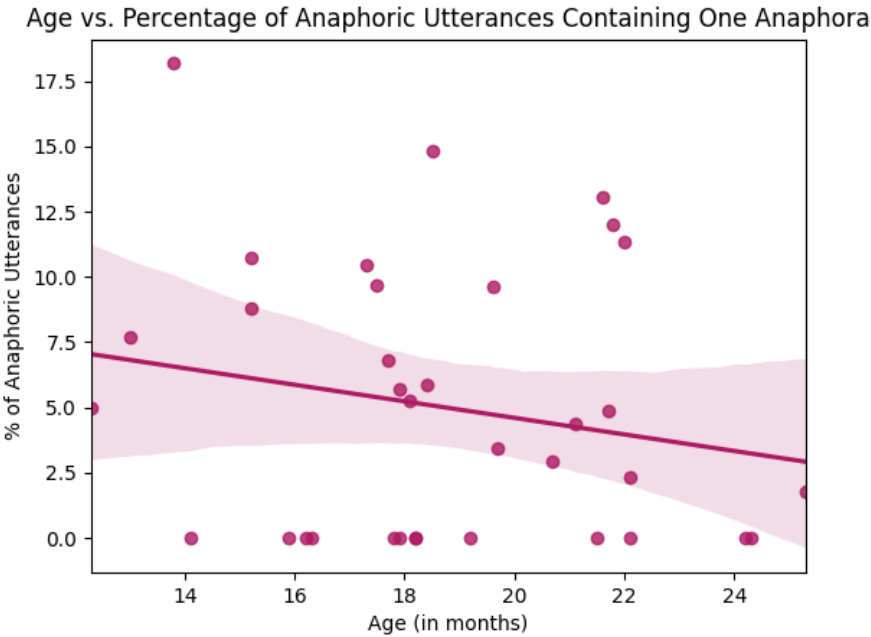


Figure 9: Compared to child age. $r(33) = -.199, p = .252$

MCDI ProdTotal vs. Percentage of Anaphoric Utterances Containing One Anaphora

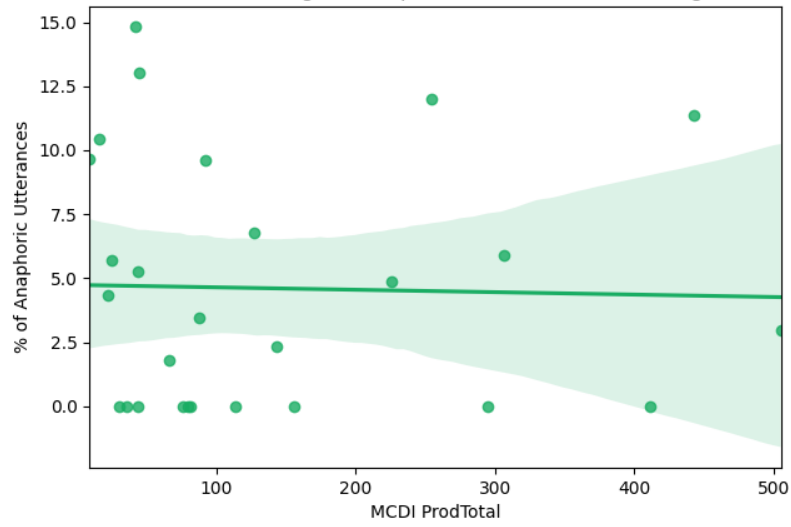


Figure 10: Compared to MCDI ProdTotal scores. $r(25) = -.028$, $p = .891$.

MCDI ProdCountNoun vs. Percentage of Anaphoric Utterances Containing One Anaphora

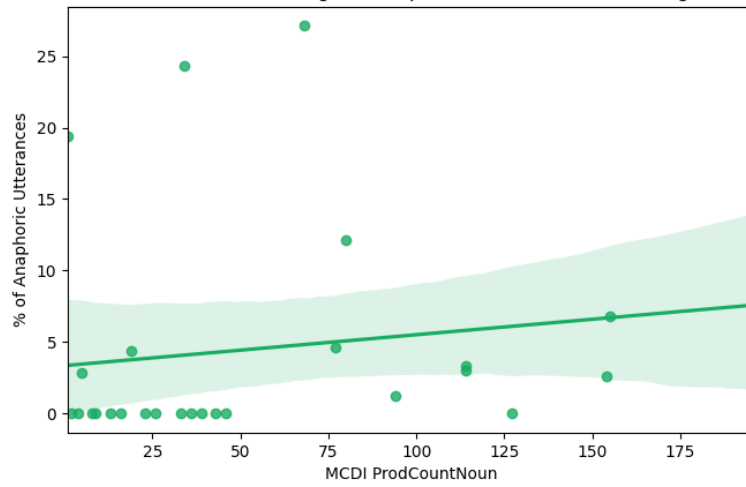


Figure 11: Compared to MCDI ProdCountNoun scores. $r(25) = -.053$, $p = .790$.

3.6 Percentage of anaphoric utterances containing verbally-cued anaphora

mean	57.86
stdev	18.19
median	60.00
max	93.55
min	22.22

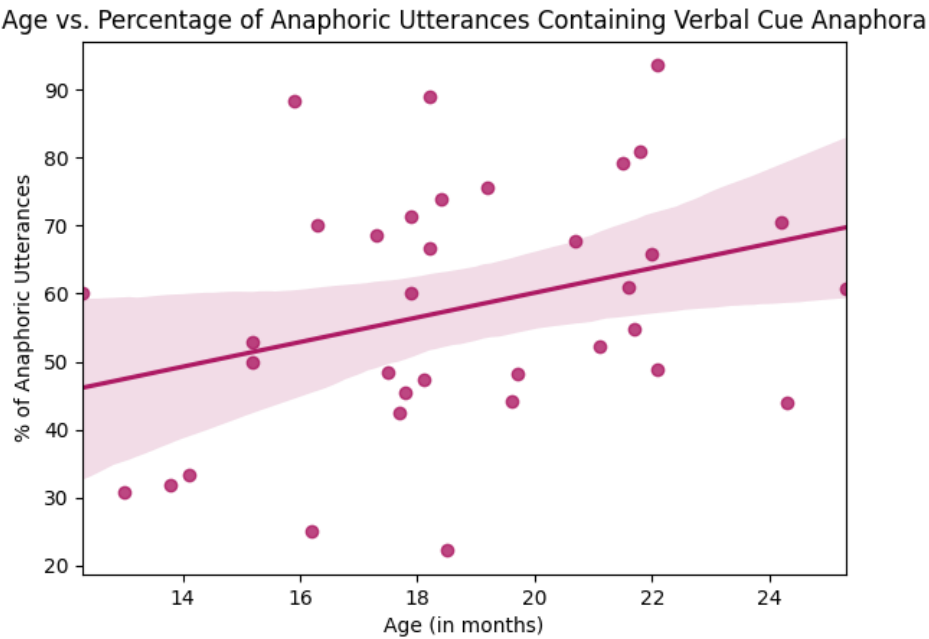


Figure 12: Compared to child age. $r(33) = .322, p = .0589$.

MCDI ProdTotal vs. Percentage of Anaphoric Utterances Containing Verbal Cue Anaphora

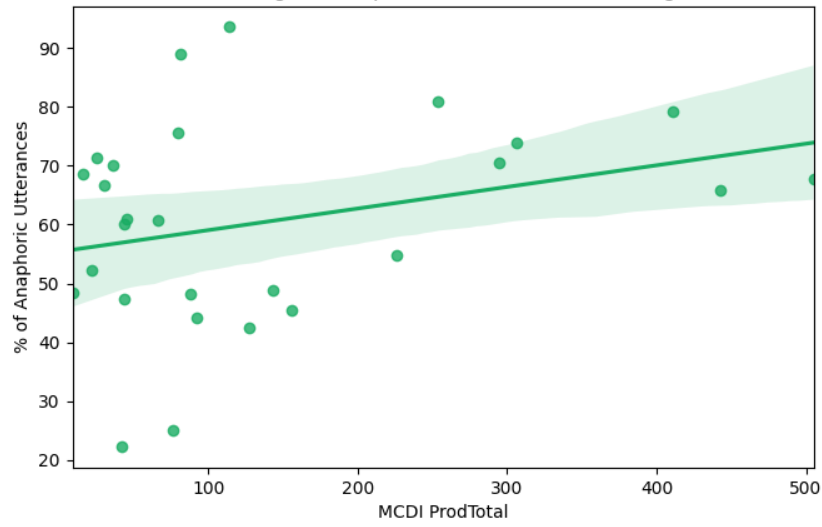


Figure 13: Compared to MCDI ProdTotal scores. $r(25) = .297$, $p = .132$.

MCDI ProdCountNoun vs. Percentage of Anaphoric Utterances Containing Verbal Cue Anaphora

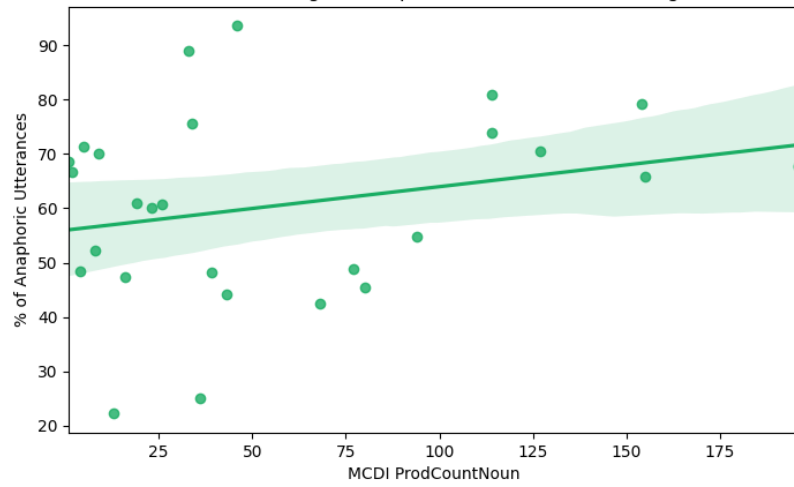


Figure 14: Compared to MCDI ProdCountNoun scores. $r(25) = .253$, $p = .203$.

3.7 Percentage of anaphoric utterances containing visually-cued anaphora

mean	43.02
stdev	18.03
median	40.00
max	77.78
min	6.45

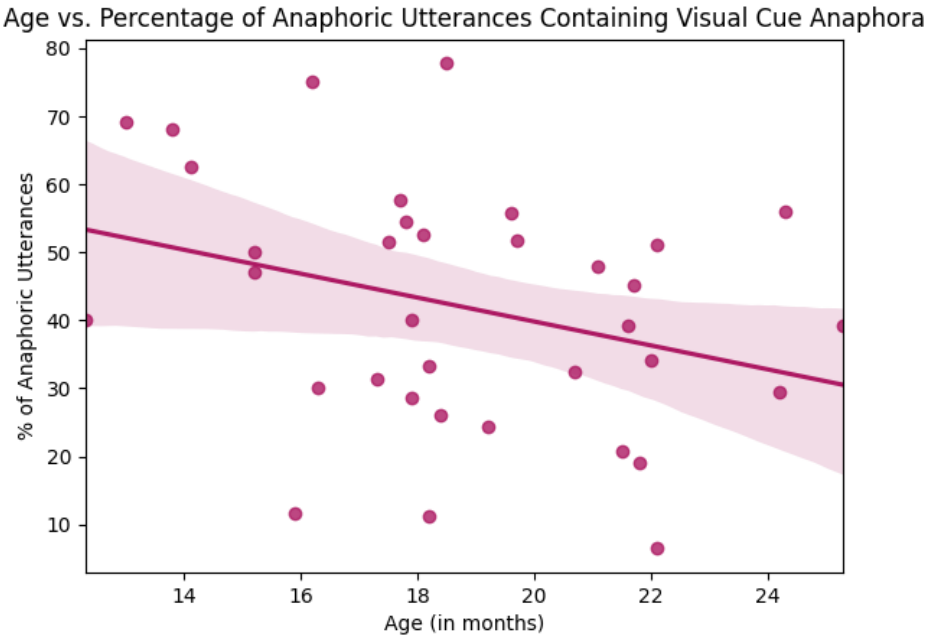


Figure 15: Compared to child age. $r(33) = -.315$, $p = .065$.

MCDI ProdTotal vs. Percentage of Anaphoric Utterances Containing Visual Cue Anaphora

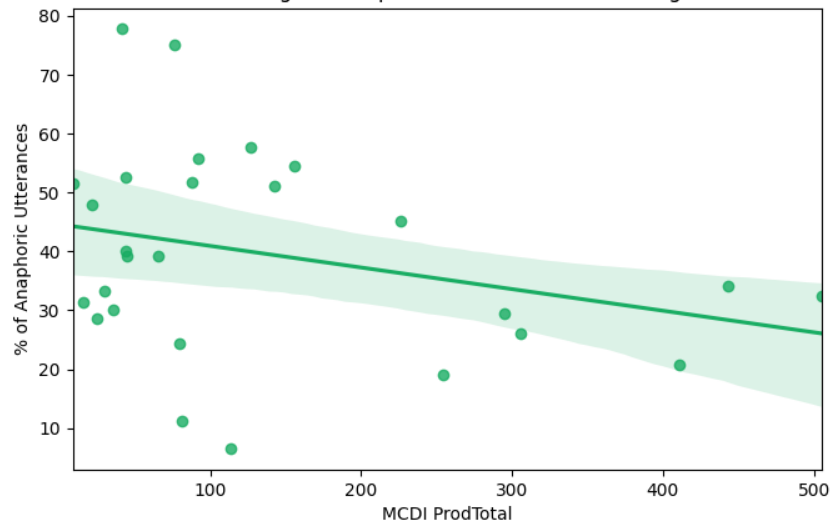


Figure 16: Compared to MCDI ProdTotal scores. $r(25) = -.297$, $p = .132$.

MCDI ProdCountNoun vs. Percentage of Anaphoric Utterances Containing Visual Cue Anaphora

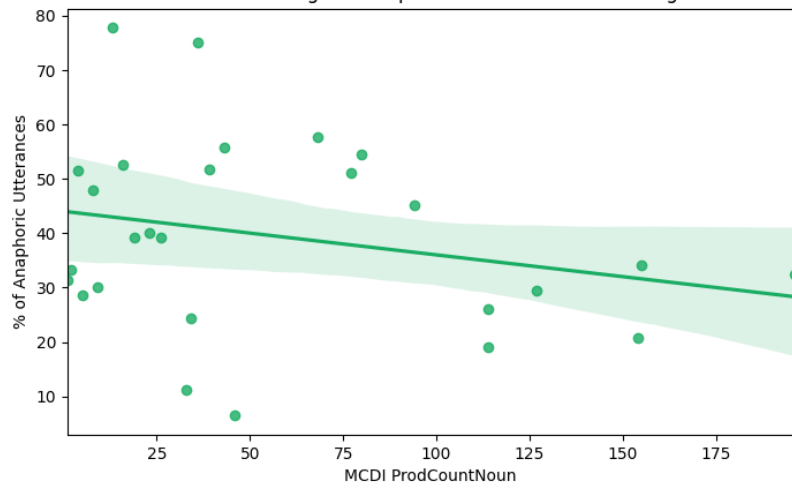


Figure 17: Compared to MCDI ProdCountNoun scores. $r(25) = -.253$, $p = .203$.

3.8 Anaphora resolution accuracy score

mean	0.618
stdev	0.144
median	0.630
max	0.903
min	0.250

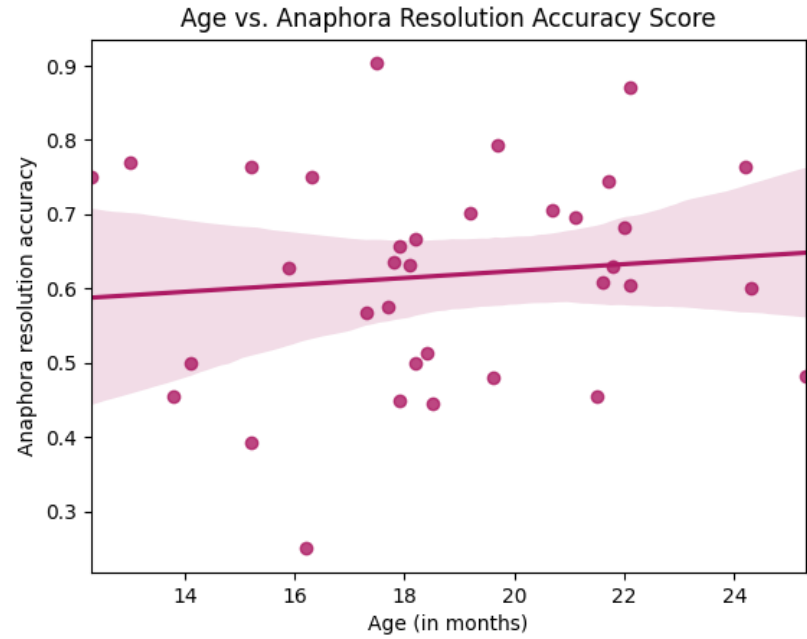


Figure 18: Compared to child age. $r(33) = -.105, p = .549$.

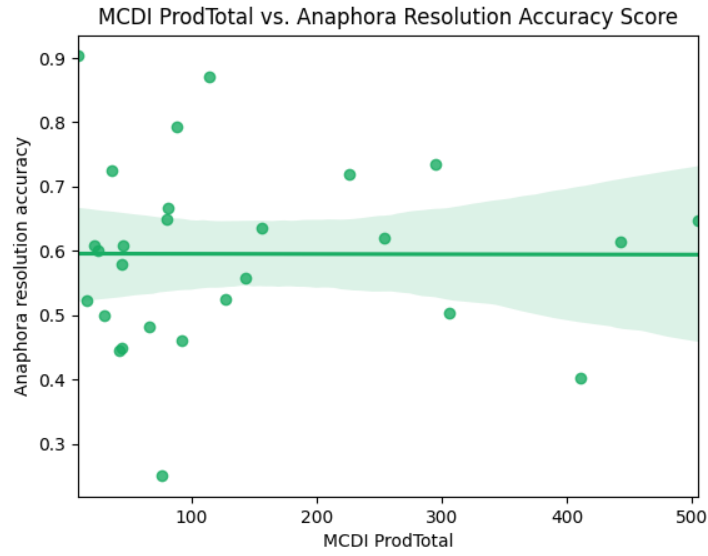


Figure 19: Compared to MCDI ProdTotal scores. $r(25) = -.003$, $p = .988$.

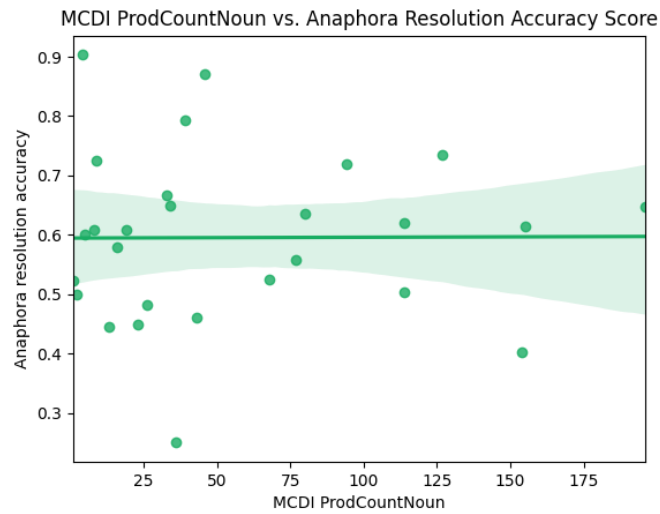


Figure 20: Compared to MCDI ProdCountNoun scores. $r(25) = .005$, $p = .980$.

3.9 Pronominal anaphora resolution accuracy score

mean	0.623
stdev	0.156
median	0.611
max	0.893
min	0.250

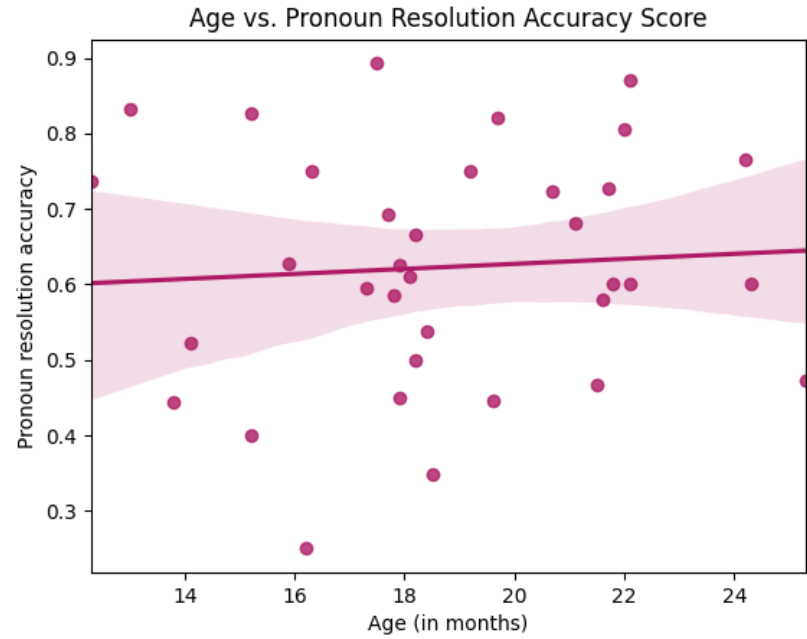


Figure 21: Compared to child age. $r(33) = 0.069$, $p = .695$.

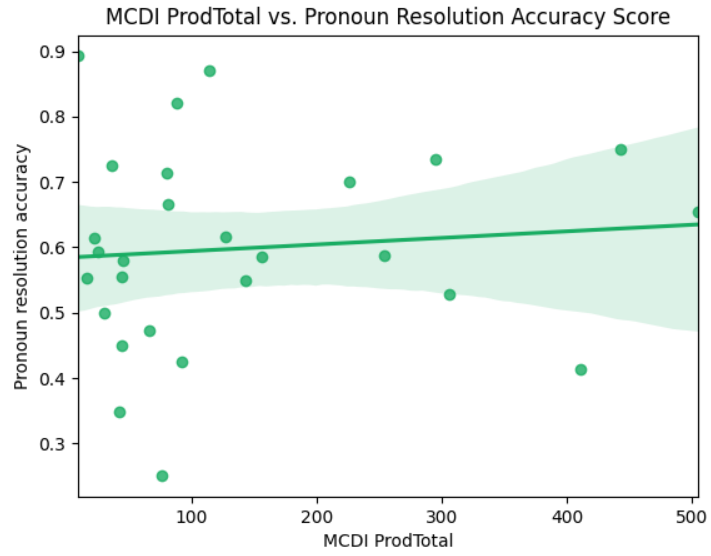


Figure 22: Compared to MCDI ProdTotal scores. $r(25) = .092$, $p = .647$.

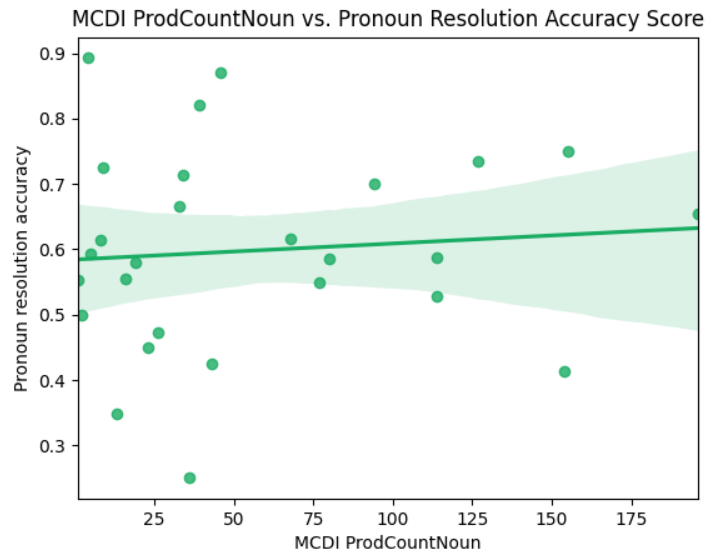


Figure 23: Compared to MCDI ProdCountNoun scores. $r(25) = .088$, $p = .663$.

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	0.707
stdev	0.356
median	0.845
max	1.000
min	0.000

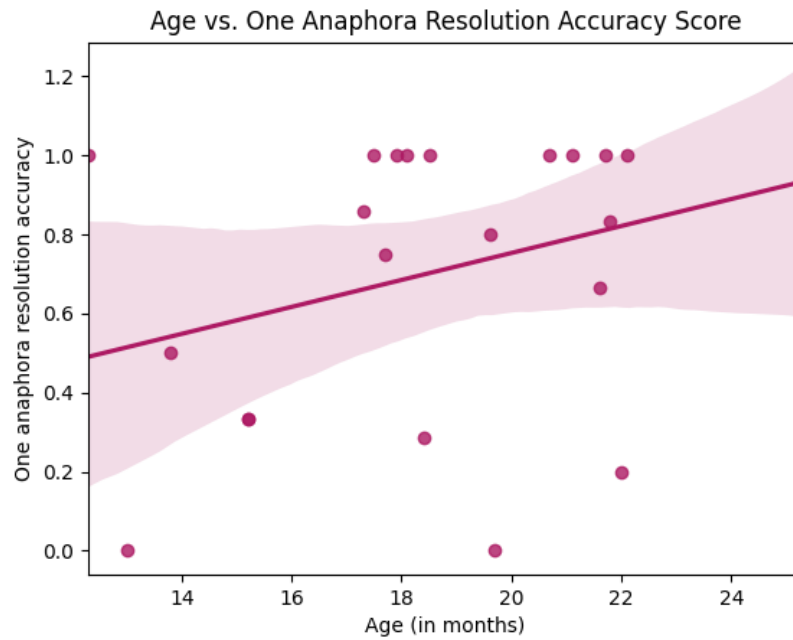


Figure 24: Compared to child age. $r(20) = .318$, $p = .149$.

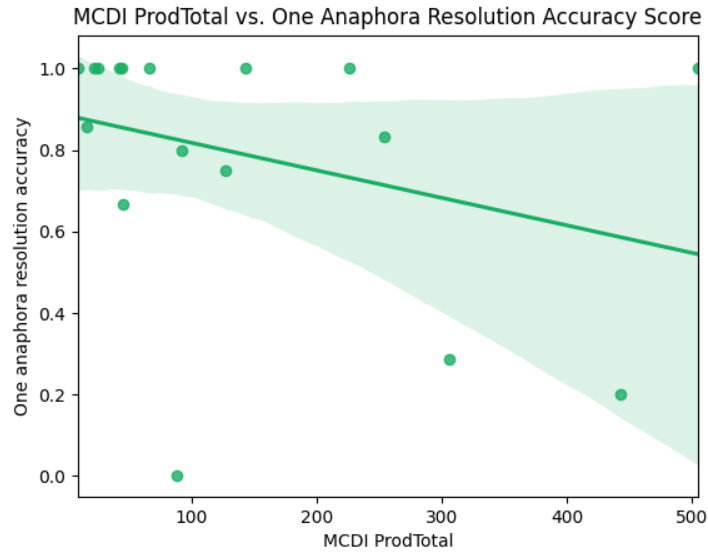


Figure 25: Compared to MCDI ProdTotal scores. $r(15) = -.320$, $p = .210$.

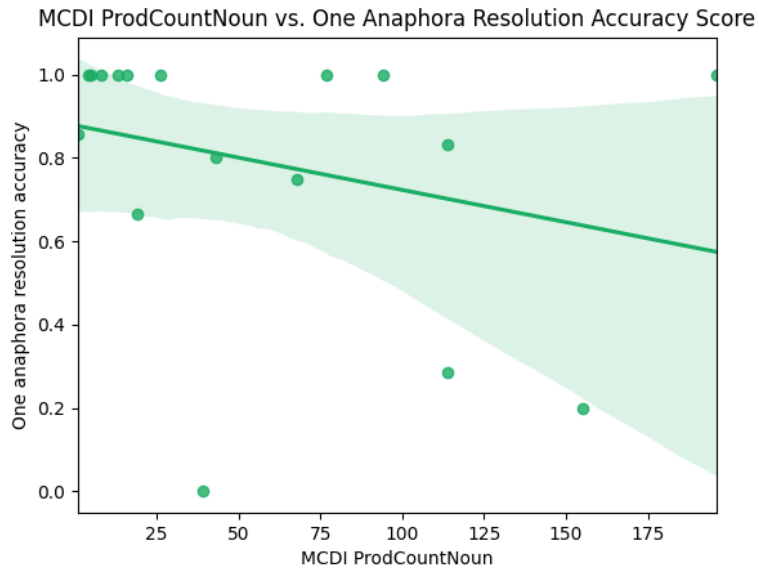


Figure 26: Compared to MCDI ProdCountNoun scores. $r(15) = -.282$, $p = .271$.

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	0.502
stdev	0.373
median	0.500
max	1.000
min	0.000

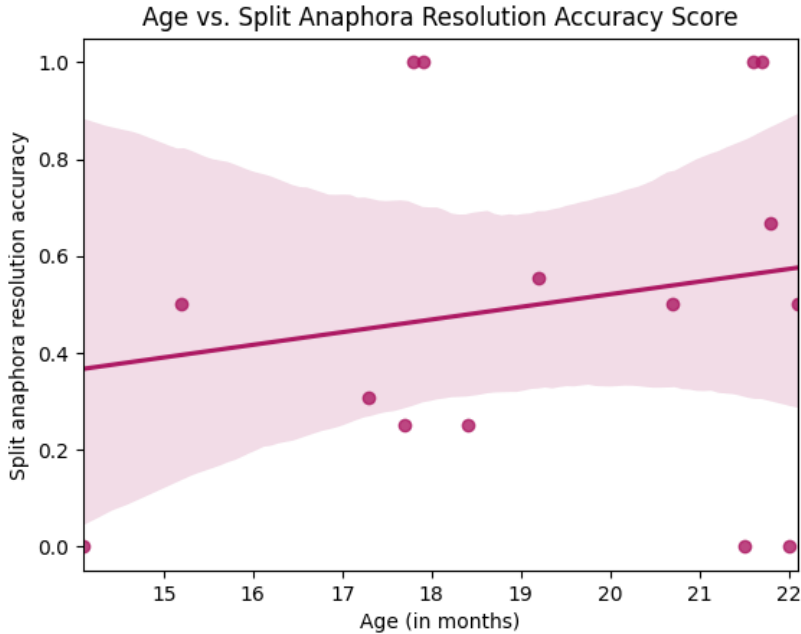


Figure 27: Compared to child age. $r(13) = .182, p = .516$.

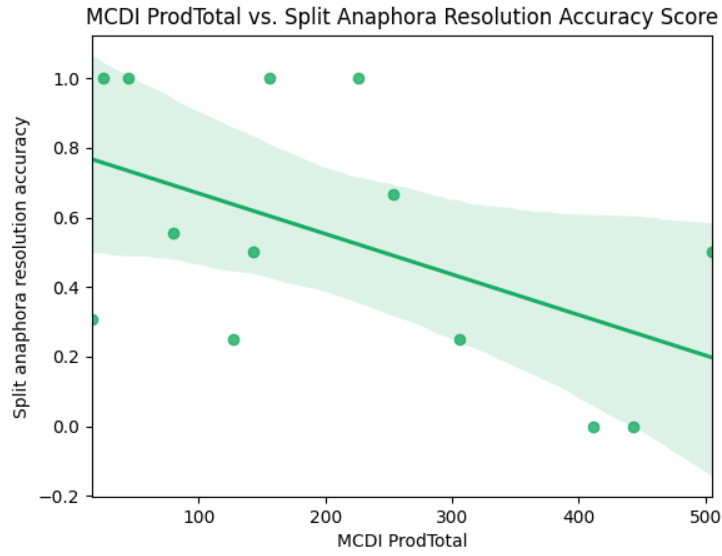


Figure 28: Compared to MCDI ProdTotal scores. $r(11) = -.512$, $p = .073$.

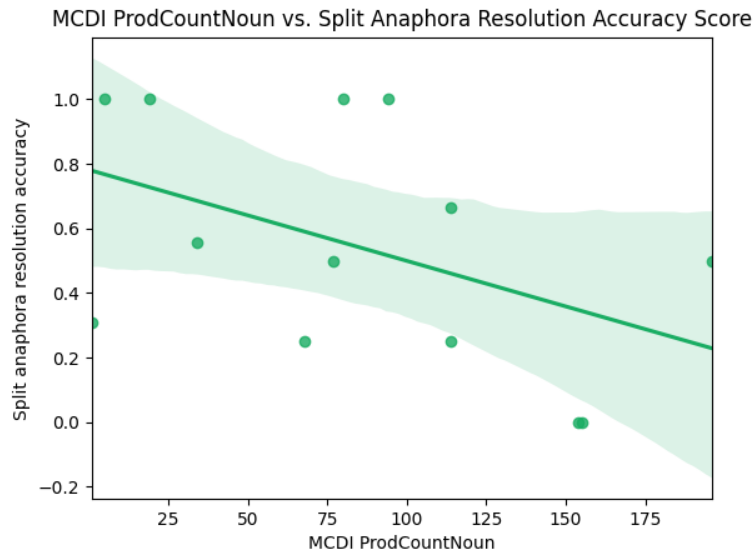


Figure 29: Compared to MCDI ProdCountNoun scores. $r(11) = -.460$, $p = .113$.

3.12 Verbally-cued anaphora resolution accuracy score

mean	0.620
stdev	0.175
median	0.625
max	1.000
min	0.261

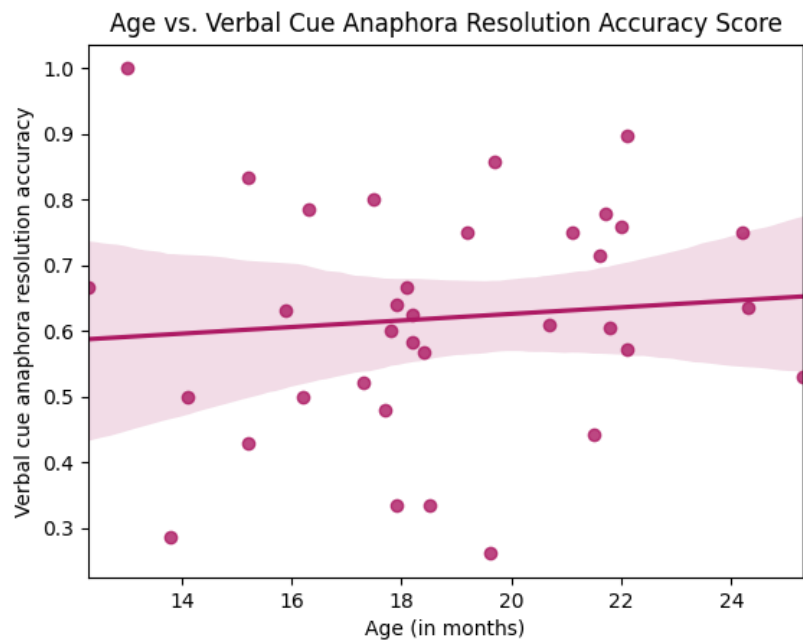


Figure 30: Compared to child age. $r(33) = .093, p = .596$.

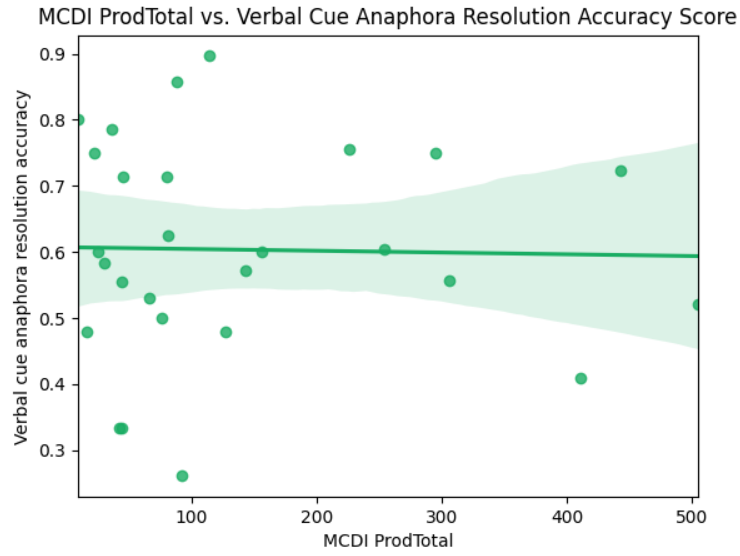


Figure 31: Compared to MCDI ProdTotal scores. $r(25) = -.023$, $p = .910$.

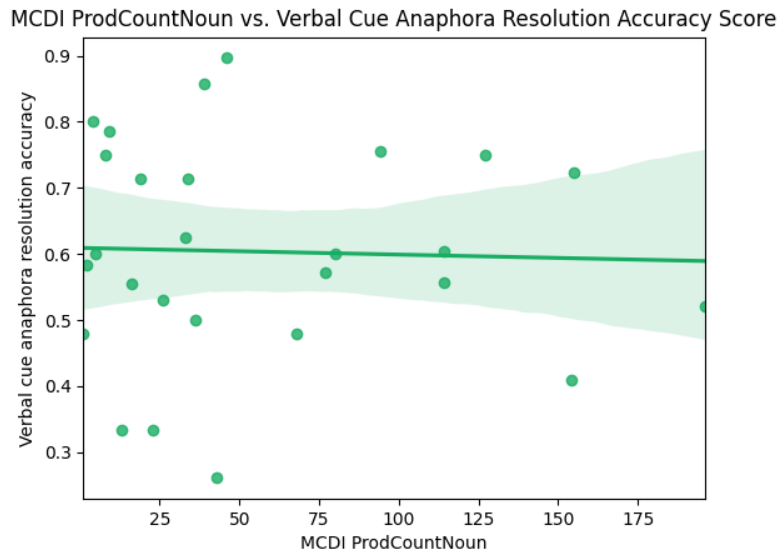


Figure 32: Compared to MCDI ProdCountNoun scores. $r(25) = -.034$, $p = .866$.

3.13 Visually-cued anaphora resolution accuracy score

mean	0.614
stdev	0.180
median	0.636
max	1.000
min	0.167

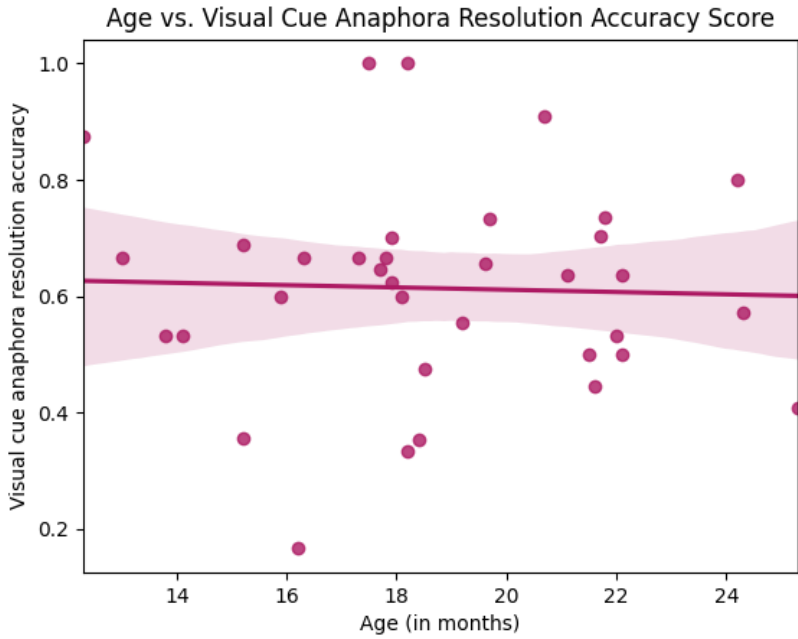


Figure 33: Compared to child age. $r(33) = -.035, p = .840$.

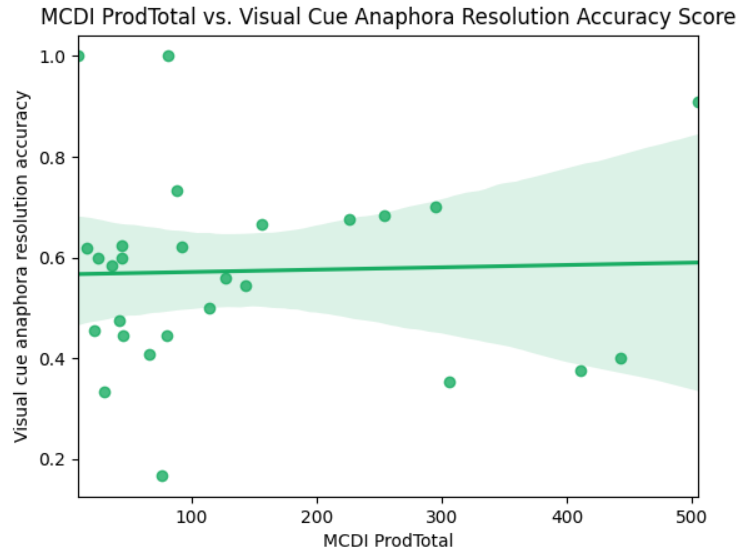


Figure 34: Compared to MCDI ProdTotal scores. $r(25) = .033$, $p = .869$.

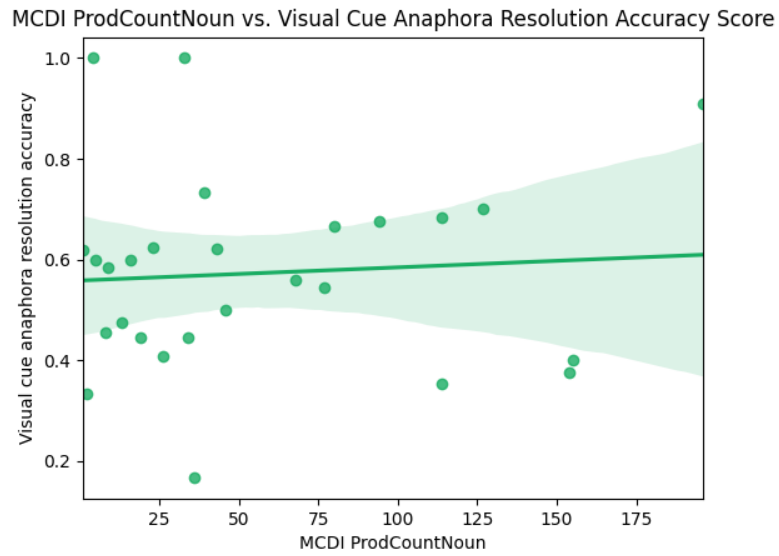


Figure 35: Compared to MCDI ProdCountNoun scores. $r(25) = .073$, $p = .717$.

4 Summary

4.1 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=0.600, new mean=0.618).

4.2 New Results & Analysis

Figures 1, 2, 3, 4, 5, 12, 15 report statistically significant findings (or almost significant in the case of Figures 12 and 15) using a significance level of .05. However, few strong correlations beyond those were found based on the reported Pearson r values.

These findings primarily support **Hypothesis 3**, that parent speech reflects the parent's perception of/assumptions about their child's language comprehension abilities. Figures 3, 4, 5 most clearly indicate this, showing that parents use more anaphora in dialogue with children who are older, as well as children who are more linguistically competent based on their MCDI scores.

Additionally, Figures 12 and 15, although not quite statistically significant, show much stronger correlations than most of the other correlations measured. This finding would support Hypothesis 3 as well. Figure 15 shows that parents used more anaphora paired with a visual cue (such as pointing to an object) for children who were younger, and fewer visual cues for older children. In other words, parents expected younger children to require greater assistance in resolving anaphora through the use of visual cues, and older children to require less. Figure 12 shows the inverse of the results in Figure 15, since verbal and visual cues were coded with binary values.

Lastly, Figures 1 and 2 indicate that there is a statistically significant positive correlation between child age and MCDI scores.

4.3 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.