Anaphora Project Update

Jasmine Falk

September 23, 2020

1 Brief Review

Hypothesis

Parents adapt their speech based on their perception of their children's language knowledge.

Child-driven anaphora

Child-driven anaphora refers to cases when a parent uses an anaphor to refer to an object that the child is already attending to. For example:

```
what did you find?
rake
car
oh it has got buttons
```

Visually-cued anaphora

Visually-cued anaphora refers to cases when a parent pairs an anaphor with a visual cue (e.g. pointing, picking up, etc.), whether or not that visual cue is required to resolve the anaphor. Previously, I had only coded for anaphora that required a visual cue to resolve it (could not be resolved from text alone).

2 Descriptive Statistics

1. **Parent speech rate:** total number of utterances per minute
2. **Anaphora utterance rate:** number of utterances containing anaphora per minute
3. **Anaphora use ratio:** anaphora utterance rate / parent speech rate
4. **Child-driven anaphora proportion:** proportion of all anaphoric utterances that are child-driven
5. **Visual cue anaphora proportion:** proportion of all anaphoric utterances that are paired with a visual cue

2.1 Overall

	mean	stdev	min	max
parent speech rate	15.168	4.721	2.904	23.179
anaphora utterance rate	4.022	2.016	0.000	8.777
anaphora use ratio	0.257	0.095	0.000	0.457
child-driven anaphora prop	0.461	0.245	0.000	0.939
visual cue anaphora prop	0.539	0.219	0.000	1.000

2.2 Data split by median MCDI score

med = 81

2.2.1 Below median score

	mean	stdev	min	max
parent speech rate	14.903	4.342	7.503	22.823
anaphora utterance rate	3.541	1.237	1.751	5.105
anaphora use ratio	0.240	0.061	0.125	0.326
child-driven anaphora prop	0.394	0.248	0.081	0.850
visual cue anaphora prop	0.616	0.225	0.286	0.947

2.2.2 Above median score

	mean	stdev	min	max
parent speech rate	17.317	3.781	11.385	23.179
anaphora utterance rate	5.149	2.365	1.368	8.777
anaphora use ratio	0.293	0.109	0.091	0.457
child-driven anaphora prop	0.578	0.208	0.190	0.939
visual cue anaphora prop	0.516	0.216	0.061	1.000

2.3 Data split by median age

med = 18.2 months

2.3.1 Below median age

	mean	stdev	min	max
parent speech rate	14.823	5.388	2.904	22.823
anaphora utterance rate	3.433	1.420	0.000	5.780
anaphora use ratio	0.227	0.087	0.000	0.348
child-driven anaphora prop	0.458	0.285	0.000	0.939
visual cue anaphora prop	0.532	0.272	0.000	0.947

2.3.2 Above median age

	mean	stdev	min	max
parent speech rate	15.450	4.282	7.503	23.179
anaphora utterance rate	4.518	2.408	1.368	8.777
anaphora use ratio	0.282	0.099	0.091	0.457
child-driven anaphora prop	0.475	0.212	0.081	0.839
visual cue anaphora prop	0.557	0.162	0.308	1.000

3 Coding Protocol

3.1 Complete Procedure

1. After you open the speech transcription file, insert **SIX** columns between **Column B** and **Column C**.
2. Change the file name to speech_#####-coded.txt, where ##### is the participant ID.
3. Check each sentence or phase in **Column I** to see if there is an anaphor.

   a. If there is not an anaphor, leave the inserted columns blank.

   b. If there is an anaphor:

      i. In **Column C**, write the anaphor word or phrase. If there are multiple anaphora in the sentence/phrase in Column F, list all anaphora in Column C, separated by commas.

      ii. Determine what object the anaphor is referring to. In **Column D**, write the referent ID (see Section 3.2).
      - If there are multiple anaphora listed in Column C, list the reference IDs for each anaphor in Column D, separated by commas.
      - If one anaphor references multiple objects, write the IDs separated by forward slashes. For example, if "they" refers to both the ladybug and praying mantis, write: 11/12.
      - If the anaphor references an object or entity that is not currently present in the room, mark the ID as 0.

      iii. In **Column E**, mark the **type of anaphora**:

one anaphora	`one`
split anaphora	`split`
pronominal anaphora	`pronoun`

      If there are multiple types of anaphora listed in Column C, list each type of anaphora in Column E, separated by commas.

      iv. *NOTE: cue variable now deprecated due to new annotations.*
      *In **Column F**, write the **cue variable**. If the anaphora can be determined by speech only, write "1". If the anaphora needs both speech and visuals to be determined, write "2". If there are multiple anaphora listed in Column C, list the disambiguation variable for each anaphora in Column F, separated by commas.*

      v. In **Column G**, write the **instigator variable**. If the anaphoric instance was child-driven, write "1". Otherwise, write "0".

child-driven anaphora	`1`
parent-driven anaphora	`0`

      vi. In **Column H**, write the **excess cue variable**. If the parent did not use any extraverbal cues, write "1". If the parent used an extraverbal cue paired with an anaphoric expression that is *not* required to resolve the expression, write "2". If the parent used an extraverbal cue that is required to resolve the anaphoric expression, write "3".

no visual cues	`1`
non-required visual cues	`2`
required visual cues	`3`

3.2 Referent IDs

Numbers are used to identify each of the toys in study. Images of the toys can be found in the anaphora project Google Drive folder in anaphora project/toys. Each toy corresponds to one unique ID value, which will be used when coding every transcript file.

Toy Name	ID
helmet	1
house	2
blue car	3
flower	4
elephant	5
rabbit	6
snowman	7
SpongeBob block	8
turtle	9
hammer	10
ladybug	11
praying mantis	12
green car	13
saw	14
doll	15
phone	16
Rubik's Cube	17
rake	18
truck	19
white car	20
spinning ladybug drum	21
purple block	22
bed	23
beach ball block	24
people	25
non-study objects	26

Notes on special cases:

- **People:** if the child or the parent is referenced, use the ID number 25. If other people not in the room are referenced, do not code for them.
- **Non-present objects:** code for abstract concepts and entities that are not present in the room that are referenced by anaphora
- **Additional objects:** for objects that are referenced that are not explicitly part of the study but still present in the room, use the ID number 26.

3.3 Summary

Column C	anaphora expression(s)
Column D	referent ID(s)
Column E	anaphora type(s)
Column F	*cue variable(s)–deprecated*
Column G	instigator variable(s)
Column H	excess cue variable(s)