Concrete Referents Coding Manual: Language Development Project

Authors: Madeline Meyers, Nicole Burke, Daniel Yurovsky
The University of Chicago

1 Transcripts Overview

The Language Development Project is a longitudinal study that follows a diverse sample of children and their parents periodically from when they first begin to acquire language through high school. Researchers filmed parents and their children interacting for 90 minutes. This may include them making a snack, playing outside, reading books, or whatever activities the parent decides. The transcripts only include speech that was directed at or produced by the child.

The spreadsheets have the following column heads:

Subj: SubjectID

Session: Which session in the Program Project; corresponds to child age. We are working with sessions 1 (14mos), 2 (18mos), 3 (22mos), 4 (26mos), 5 (30mos), and 6 (34mos).

Time: timestamp (times are recorded for *gestures* only).

Person: Person who spoke/gestured (child/parent). Parent includes either mom or dad, and does not specify either.

Chat: Utterance spoken–contains spoken referent(s)

Form: Form of gesture–we are coding *deictic* gestures, or gestures that take the place of a noun. The deictic gestures we code for have should have the form of *hold*, *point*, continuous point (*cont point*) , *palm*, or *cont palm.*

Gloss: What did the gesture mean? For this project, we are **not** coding gestures that have *take* in their gloss. We are coding gestures that imply noun(s) or giving (a) noun(s), not gestures that refer to taking nouns or to locations in space. More on this below.

Obj: Gestured referent

Gs_rel: What function did the gesture serve? We are currently not using this column but may in the future.

Context: Provides extra information about the environment.

The following columns are what you are going to fill in from the transcripts. Multiple referents should be separated by a semicolon ";".

Spoken_obj: Any concrete noun from the chat column. More on this later.

Gestured_obj: Any referent from the obj column.

Referent: What was the referent? This column should have the same referents as "Spoken_obj" and "Gestured_obj" combined. In some cases, there will be more than one referent.

Comments: A place to mark if anything is weird/you're unsure about and you want a Senior Researcher to look at it.

Ref_pres: If the video has been watched for present referents, whether the referent is present (1=yes, 0=no) will be included here. Most files do not have this column, only the predicted column. The order of numbers in these cells correspond to the order of referents in the referent column **Ref_predicted:** Do we think the concrete referent is present at the moment? Would the person be able to gesture to the referent? 1=yes, 0=no.

Coding Guide

2   Goals

We are interested in coding all the concrete nouns spoken throughout the session by both parent and child. They may not necessarily be present at the moment, but they are nouns that can be gestured to if present. Generally speaking, present nouns will be used for the gesture/speech tradeoff portion of this project, and all nouns will be used for the discourse/topic portion of this project.

3   Coding A Referent

When deciding which referents are concrete, ask yourself if they can be pointed to, regardless of whether you believe they are present in the environment. This is the most useful determiner for what should be included.

- Referents should be converted to their **lemmas**. A lemma is the "dictionary form"–usually the singular form–of the word. Here are some examples of lemmas:

  - teeth = tooth
  - stairs = stair
  - pajamas = pajamas
  - doggie = dog
  - horseys = horse

- Each referent should be in the form of **one word**–if it is a two-word phrase, or a noun that is more specific, connect the words with an underscore "_"

  - birthday cake = birthday_cake
  - belly button = belly_button
  - Christmas tree = christmas_tree

- Multiple referents should be separated with a **semicolon** ";"

- Referents should be coded using **lowercase** letters only

- Figure out what the child and siblings' names are so you can code them as necessary

4 Guidelines

4.1 Defining A Referent

What is a concrete noun? It depends on the context. We want to capture any nouns that are gestured to or that are spoken in the transcript, which includes:

- **Proper nouns** that could be found across subjects

  - All **family members**, even if it is the actual name of the family member that is said. For example, all references to the child's name will be coded as "child", or if you know that a child's sibling is named Esmerelda, all mentions of esmerelda should be coded as "sister."

  - Additionally, if you know the name of a **pet**, code references to the pet as whatever animal the pet is (ex. if the family has a cat named Smelly, all references of Smelly the cat should be coded as "cat").

  - **Common characters** included in the following list. These characters were included because they were present across subjects and across ages. Therefore, they functioned as any other concrete noun–they did not represent just a single object, but were found across media in different forms. All other character names should be coded as what they actually represent if possible (ex. minnie mouse is a mouse, cinderella is a princess, spiderman is a superhero, bob the builder is a construction_worker, nemo is a fish). More obscure references that can't be generalized (ex. Goldilocks is just a girl), can be left uncoded. Similarly, we do not code everything as "toy."

    Characters that should be coded
    - Pooh = pooh
    - Tigger = tigger
    - Elmo = elmo
    - Mickey = mickey_mouse
    - Scooby Doo = scooby
    - Spongebob = spongebob
    - Cookie Monster = cookie_monster
    - Patrick (from spongebob) = patrick_star
    - Barney = barney
    - Dora = dora
    - Any Teletubby character (Tinkie Winkie, Dipsie, La la, or Poe) = teletubby
    - Any Star Wars character = star_wars
    - Santa Claus = santa
    - Rescue Hero toy = rescue_hero

- Universal **rooms** in the house

  - ex. different rooms like living_room, bedroom, kitchen, bathroom

- **Shape words**, including smiley_face (:, if they are referring only to the shape and are not background words

- **Nicknames for objects**, as long as you are certain of what object the nickname refers to. For example, if you know that "bidiba" is the child's name for blanket, every time the child says "bidiba" you should code it as "blanket"

Our definition of a concrete noun does **not** include:

- **Pet names** (ex. do not code "honey" as a referent for the utterance "Oh honey, it's okay")

- Words that have **multiple meanings** (ex. if the child is saying "cheese" to the camera, do not code that as a referent for cheese, or if someone says "you're chicken", do not code that as a referent for chicken)

- **Baby** animals, such as "puppy" should be coded in their adult form, "dog", unless the animals are being compared to one another (ex. there is a story about a mom dog and a puppy)

- **Background objects** including words like sky, page, paper only coded when stand alone, not if there is a foreground referent present

- **Pronouns**, like "me" and "you" are not coded as concrete referents for the people they refer to

- Words that **could be nouns but are being used as verbs**, even if the usage is switching rapidly

  - Ex. "Hammer the worm"(V) versus "get your hammer"(N) or "button that"(V) versus "how many buttons?"(N) "Go potty"(V) versus "Go to the potty"(N)

- Gestures to **vague locations**

  - ex. a gesture to an object that is "over there" or "used to be there"
  - general terms like "bottom", "top", "back", "front", "inside", "outside", although "lid" is a concrete referent

- **Gestures that cannot be identified**. Typically, all deictic palm/point/hold/give gestures will be coded for, except for those where the Obj is undefined

- **Numbers**

- References to the experimenter

5 Potential Areas of Confusion

5.1 Color

Sometimes, parents and children will speak about the color of the object instead of the object itself, but they will also point to the object. Because it is difficult to distinguish the intention of their point, code all gestures as referring to the object that is being pointed to, and only code verbal referents if they include a concrete noun. For example, if a child says "Blue" and points to a blue crayon, you should leave spoken_obj blank and code "crayon" for gestured_obj.

5.2 Siblings

It is sometimes difficult to know if a sibling is present or not. Try watching a few minutes of the video to get a feel for the layout of the environment and persons present in said environment.

5.3 Photos

In photos, it is difficult to distinguish what is foreground, what is background, and who is being mentioned. The transcripts are generally coded very well to indicate that a gesture is toward a particular character (ex. grandma) in the photo. If you are not able to identify the relation of a person (ex. Shelby–who is she?), leave spoken_obj blank, and indicate that the gestured referent is "photo."

5.4 Nouns vs. Verbs

If you are unsure whether a word is being used as a noun or verb, and cannot tell based on context clues, as often happens with child speech, take the conservative route and assume the referent is being used as a verb, because that will work against our hypotheses.

5.5 Counting

Sometimes, when parents and children are counting, they will say the numbers while pointing to objects. Just as with colors, code the gestured objects but do not code the numbers under spoken_obj.

5.6 Letters

If either the gestured or spoken columns contain the word "letter", code the referent as "letter". Do not code a referent of child if the child's name is spelled out letter-by-letter. All letters that are said individually have an "@" next to them, so they are easy to distinguish.

5.7 One vs. Multiple Referents

There will be situations where you are unsure whether to code an object as a single referent using an underscore, or as multiple separate referents. As a rule of thumb, if the whole is not the sum of the parts, and the referent is common, code it as one word. If the whole is the sum of the parts, especially if the referent is uncommon, code it as two separate referents. For example, when you see the words "ice cream", you have an immediate mental image of a lovely dessert. Seeing the words "ice cream" do not activate your "ice" or "cream" schemas, because "ice cream" is its own separate category, which is just as large or larger than "ice" or "cream" categories. Therefore, the whole is not the sum of the parts and "ice cream" should be coded as one word (ice_cream). However, when you hear the words "mushroom soup", there are multiple schemas activated. You probably think of different kinds of mushrooms as well as different kinds of soups. In this instance, the whole is the sum of the parts, because mushroom soup is a member of the larger soup category and a member of the smaller mushroom category, but is not a category on its own.

5.8 Pretend Play

Pretend play is most confusing when coding gestures and referent presentness. You should take what the child says as the rule. For example, if a child creates a blob-like playdoh creation and says "that is a snake", the gestured_obj should be coded as snake (not playdoh; snake, as playdoh is the background feature in this context). If the parent counters and points to the pretend snake and says, "that is a rope", you should code her gestured_obj as a rope, even though the snake and the rope are the same object.

5.9 Referent Specificity

How specific should a referent be? It depends on the context. A referent should always be coded as the most specific way it is referred to in the surrounding discourse, even if it is mentioned at different levels of specificity. A referent should not be coded as being more specific than it is, if that specificity is never mentioned in the discourse. There is only one exception to this rule–if the less specific referent has a contradictory meaning from the actual referent. For example, if the parent and child refer to a wood chip on the ground only as a "chip," it should still be coded as a wood_chip. In these cases, there is usually a more specific referent mentioned in the context or Obj columns. Here are some general guidelines:

- stuffed animals, real animals, and pictures of animals are coded the same (as the animal) *except* teddy bears because they are mentioned widely across children and seem to be a unique category (coded as teddy_bear)

- similarly, baby dolls are coded as babies and not both babies and dolls

- pieces of objects are coded separately if they are referred to or gestured to multiple times in the discourse (ex. puzzle_piece vs. puzzle)

- fake/toy objects and real objects are coded the same (as the object)

- if you know that a parent and child are talking about a toolbelt, as they use the word "toolbelt" earlier in the discourse, but then switch to using the word "belt", continue to code "belt" as "toolbelt".

- generally speaking, occupations are coded as one referent rather than two, for example, an ice cream man would be coded as ice_cream_man instead of ice_cream; man, because this second package does not contain the whole meaning of an ice cream man.

5.10 Synonyms

A lot of the time, you will run across words that have the same meaning, and we want to code those as the same referent! Check out the handy dandy "Quick Coding Vocabulary Chart" at the end of the document to see what vocab to use.

6 Predicting if the Referent is Present

Not all concrete referents that are mentioned in speech are actually present in the environment, because people commonly talk about objects that are in other rooms, things that happened in the past, or things that have yet to happen. However, it is necessary to identify which referents are in the environment, especially because it is impossible to gesture to an object that is not present. Therefore, the goal of determining referent presentness is to make sure that all referents included in the speech/gesture tradeoff analysis have nonzero probabilities of being spoken about or gestured to.

While it would be ideal to watch every video to determine whether the referent is present in the environment, that is not feasible, so we will be predicting whether the referent is present by using information in the transcripts.

It is helpful to watch the beginning of every video to understand who is present, and what the layout of the environment is. Can you point to the kitchen from the living room? The bedroom

from the bathroom?

Keep in mind that objects can be present in the environment without the subjects attending to the object. It is common that one person sees an object, but the other does not. Coding these predictions takes great attention to detail.

Context clues are enormously helpful. If an object is gestured to at any point while the parent and child are in the same location, this object can be assumed to be present in the environment. However, if either the parent or the child does not know where the object is (ex. "where is x?"), they could not have pointed to the object, so it should be counted as not present.

Objects that are present are coded as "1", and objects that are not present are coded as "0". You should include one code for every referent in the utterance, so if there are multiple referents, use semicolons to separate your 1's and 0's.

It is necessary to consider whether the referent is not just present, but pointable, and how a gesture to that referent would be coded during transcription. For example, parents talk about how the batteries in a toy might be dead. Even though the batteries are present in the environment, if the parent pointed towards the toy and said "the [point] are dead", that would be interpreted as "the [point to toy] are dead", meaning that the toy is dead, instead of "the [point to batteries] are dead", which is the actual meaning of the phrase. This rarely happens with words other than battery and diaper.

If the subject believes that the referent is present, it should be coded as present. This topic only becomes confusing during pretend play, where a child makes up a new object, or changes the meaning of an item. If the child labels the item, it is present. During these pretend play situations, parents often ask what an item is, and any time the parent is not sure what the item is, the parent's referents should be marked as not present. For example, a parent may ask "Is it a lemur? Is it a monkey?" If the child says "It is a lemur," the parent's utterances should still be coded as not present for the following reason. The child could have easily said "No, it is an orangutan," in which case, the parent's utterances would have been not present as well.

One last thing to keep in mind is that the same referent can be present in one line and not present in the next depending on its meaning. For example, if a parent and child are looking at a photograph of grandmother mary, grandmother is present. If, however, in the next line, the child asks where grandmother kathleen is, grandmother is not present. This is similar to an example where there are multiple objects and only one is present–if the parent has a red ball, but the child is looking for the yellow ball.

Referents that are always present

- objects that are gestured to

- objects that have been previously gestured to unless there has been a change in room location

- body parts of people/dolls that are present

- objects that are directly labeled, ex. "this/that is..." even if they are incorrectly labeled

- the subjects that are present, though be careful to not code "Dad" as being present if he is not there. Occasionally, "mom" will not be present if it is referring to someone else's mom (ex. if parent and child are reading a story and "his mom is not there").

- all referents during storytimes are assumed to be present in the pictures of the book

Referents that are most likely not present

- "go get your X" if the target referent is in a different room

- "your Y is not here"

- "that's not a Z"

- "that goes with another X, not this one" or phrases that are the opposite of a label for the referent that is coded

- "I need..."

- "make a..." watch out for these phrases during craft projects. You cannot point to a circle that has not been drawn yet.

- "where is the X?" look at what the response is to determine if it is present or not. Parents often will be looking through books, in which case they believe the object is present, and are just quizzing the child.

- "Is it 'x' or 'y'?" There is no way the person uttering this statement could point to the specific referent of the object they are asking about if they are not sure what the object is. This happens often in pretend play, where a parent is asking what the child believes an object to be.

- things that are heard, but not seen

- rote songs or phrases

- conversations that involve stories about the past or future

7   Coding Examples

	person	chat	spoken_obj
	parent	here .	NA
	child	honey .	honey
	parent	honey .	NA
1.	parent	who's my honey ?	NA
	child	Pooh .	pooh
	parent	oh, Pooh likes honey .	pooh; honey
	parent	he does .	NA
	parent	does Pooh eat honey ?	pooh; honey

	person	chat	spoken obj	ref_pres_predicted
	parent	want a bite of banana ?	banana	1
2.	child	orange .	orange	0
	parent	that's all the orange we have .	orange	0
	parent	that's it .		

person	chat	spoken_obj	ref_pres_predicted
parent	oh CHILD, what was that ?	child	1
parent	did you see the semi ?	semitrailer	0
parent	did you see the semitrailer ?	semitrailer	0
parent	big one .		
child	small .		
parent	small ?		
child	train .	train	0
child	train .	train	0
child	train .	train	0
parent	have to wait for the train .	train	0

3.

person	chat	spoken_obj	ref_pres_predicted
parent	but the, the top to the pineapple is gone .	lid; pineapple	0; 1
parent	where's the top to the pineapple ?	lid; pineapple	0;1
parent	what+about the bananas ?	banana	1
parent	oh, here's the top to the pineapple .	lid; pineapple	1; 1

4.

person	chat	spoken_obj	ref_pres_predicted
child	I want a hat .	hat	0
child	I want a hat .	hat	0
child	want a hat .	hat	0
parent	okay, it's all ready for you .		
child	it not a hat .	hat	0
child	not a hat .	hat	0
child	no it not a hat .	hat	0
child	it not a hat .	hat	0
parent	it's not a hat ?	hat	1
child	no, it is a bunny .	bunny	1
child	not a bunny .	bunny	0
child	it is a hat .	hat	1
child	it is not a blue bunny .	bunny	0
child	it a hat .	hat	1
child	it not a blue bunny .	bunny	0

5.

person	chat	form	gloss	obj	spoken_obj	gestured_obj	ref_pres_predicted
parent	what's that ?						
parent	an eye+ball ?				eye; ball		1; 1
parent	yeah .						
child							
parent	going to juggle ?						
parent	want to juggle ?						
child							
parent		hold	give balls	balls		ball	1

6.

9

8   Quick Coding Vocabulary Chart

Word	Coded Referent
trash, trash can	garbage
stomach, belly	tummy
lace	shoelace
plane, airplane	airplane
remote control, clicker, remote	remote
rabbit, bunny, bunny_rabbit	bunny
mickey, mickey_mouse	mickey_mouse
grandma, nana, granny, etc.	grandmother
grampy, gramps, grandpa, etc.	grandfather
potty, toilet	potty
bottom, butt, tushy, tushie	butt
boo+boo, ouchie	cut
pull+up, diaper	diaper
socket, outlet	outlet
airplane, plane	airplane
Santa Claus	santa
teeter+totter	seesaw
smiley face image	smiley_face
ice cream cone, ice cream dish, etc.	ice_cream
possum, opossum	possum
tissue, kleenex	tissue
bookbag, backpack	backpack
road, street	street
stool, stepstool	stepstool
play dough, play doh, playdo, play_dough, etc.	play_doh
mower, lawn mower	lawn_mower
glass, cup, sippy cup	cup
sucker, lolipop	lolipop
noggin, head	head
mama, ma+ma, mommy	mom
dada, da+da, dad, daddy	dad
VCR	vcr
picture (=photo)	photo
picture (=drawing)	picture
Mr. Potatohead	potatohead
telephone, cell phone, landline phone	phone
fire engine, firetruck	firetruck
toy box	toybox
hippopotamus	hippo
dino	dinosaur
television	tv
drawing, painted picture	picture
photograph, "picture"	photo

refrigerator, fridge	refrigerator
clotheshanger, coathanger, hanger	coathanger
hot tub	hot_tub
donut	doughnut