

NYU-Emerson Corpus

Transcription Procedure:

Audio files were first transcribed by an external agency. Given that audio-only transcription of parent-child interaction is challenging and that automated systems often miss important details, subsequent transcription phases were crucial for providing a high quality transcript for language analyses. Therefore, each initial transcription was subjected to a series of additional steps; the transcription work flow is summarized below:

*headphones were used for the first two transcription phases

- Prepping Phase:
 - Research Assistants (RAs; who have complete training for independent prepping after reformatting at least two training transcripts and reviewing detailed feedback from lab staff) converted the audio-only transcript to one that structurally resembled a CHAT transcript removing names to de-identify.
 - RAs made edits to the transcript while watching the parent-child interaction video, improving accuracy for speaker tiers, utterance boundaries, lexical material, and untranscribed sections. This resulted in a *prepped transcript*.
- CHAT Phase:
 - A CHAT trained RA or SLP (Speech Language Pathologist; reliable via training standards set by the lab/project protocol including reaching 90% reliability for utterance boundaries and utterance delimiters on at least two training transcripts) watched the parent-child interaction and used the prepped transcript to generate a *CHAT transcript* by further improving utterance boundaries, lexical material, untranscribed or unintelligible sections, and adding specific CHAT conventions.
 - Each transcript-video set was viewed by at least two team members, always including an SLP.
- Proofreading Phase:
 - CLAN features including: CHECK, FREQ, and MOR (10.21.25 version) were run on transcripts to identify typos and unrecognized lexical material.
 - Unrecognized words were added to the corpus lexicon in a cut file (detailed below). Where the MOR lexicon has multiple common spellings for a word, we selected one spelling for internal consistency.
 - Batchalign2 was run to identify any CHAT specific typos.
 - RAs then read the transcript (no video viewing) looking for typos, mis-matched speaker tiers, etc. to identify any remaining errors before flagging it as ready for upload to TalkBank.

Transcription Conventions

Time Stamps:

- Time stamps and Headers were included for each section corresponding to the parent-child interaction protocol, and GEM Markers (@G) were added.
 - Section names include: Free play/Conversation, Pretend Play Independent, Pretend Play with Parent, Sensory/Cause and Effect Play.

- Judgement was made as to when the transcribers felt that each specific section truly started, but always after directions were provided by the INV.
- An end timestamp was added when the play finished and transcription stopped.
- The accompanying videos may start before the transcript and/or may end after the transcript based on when the section was judged to have begun.
 - All videos were trimmed following the completion of the parent-child interaction.

Privacy:

- Personally identifiable information was removed with video blurring and audio editing as needed.
- We also used the following anonymization conventions in the transcripts:
 - Use of www with an explanation tier when personal information was produced.
 - Use of Childname, Mothername, Fathename, Investigatorname, Firstname, Schoolname, Petname, etc.

Detailed CHAT Conventions:

Lexicon

- Project staff built a lexicon using CLAN’s ENG folder and MOR 10.21.25 version with additions specific to this study, including:

aac	{{scat n}}	n-abbrev
aba	{{scat n}}	n-abbrev
atv	{{scat n}}	n-abbrev
crisscross_applesauce	{{scat co}}	co-rhyme
off+roading	{{scat n}[comp n+part]}	n+n+part
parkour	{{scat n}}	n
polyhedron	{{scat n}}	n
sriracha	{{scat n}}	n

Yes/No Spellings

- mhm (meaning yes)
- uhhuh (meaning yes)
- hmhm (meaning no)
- hunhunh (meaning no)
- nahunh (meaning no)

Shortened Words

- The following *shortened/incomplete words* were transcribed as full words for readability:
 - cause→because
 - em→them
 - whatcha → what are you or what you
 - dontcha → don’t you
 - lemme → let me
- The following *shortened words* were transcribed as they were spoken:

- wanna
- gonna

Proper Nouns & Titles

- Character names and/or titles were transcribed as proper nouns (e.g. McQueen, Baby) and received an underscore as a linkage as needed (eg. Peppa_Pig, Spider_Man, Ninja_Turtles, Old_MacDonald).
- Nicknames and/or terms of endearment were capitalized and thus transcribed as proper nouns (e.g. Honey, Sweetie, Sweetheart, my Love).

Questions and Recasts

- Intonation was primarily used to determine if inclusion of a ? utterance delimiter was warranted.

Vocal Actions were transcribed using &= and were not phonetically transcribed.

- Examples include but are not limited to: laughs, gasps, sighs, grunts, squeals, screams, hums.
- If these actions (e.g. laughing) occurred along with production of lexical material, the <"> [=! laughing] notation was used.
- *Audible inhalations* (i.e. not gasps) were not coded with this notation and instead were coded in a comment tier.
 - Audible inhalations that were solely part of inhaling to blow up a balloon were ignored.

Vocalizations were judged as non-word like or word-like.

- *Non-word vocalizations*: vocalization that cannot be clearly characterized as a true word approximation. It may be sensory in nature. It is likely mostly vowel only (V) or a combination of consonant-vowel (CV or VC.)
 - A phonology tier was added as often as possible, especially for CV or VC.
 - If the speaker had a toy, food/drink, or a pacifier in their mouth, the vocalization was not phonetically transcribed.
- *Word-like vocalization*: these can be considered “words” for the purposes of identifying words/utterances. This may contain “novel” jargon or scripted unintelligible language.
 - These may include sounds (syllables of at least one vowel and one consonant or repeated syllables, jargon, different syllables strung together).
 - These were coded as yyy with an accompanying phonology tier.
 - If the word-like vocalization (yyy) is in a multi-word utterance, the phonology tier includes phonetic transcription of yyy and the ellipsis symbol for all other words and/or conventions on the corresponding speaker tier.

Gestures were not the primary focus of the project and therefore were not specifically coded using &= notation.

- Gestures that seemed to convey communicative intent were included as comment tiers for all participants.

- Common gesture descriptions include: points, beckons, nods head-yes, shakes head-no, and other descriptions such as turns hands out to side, palms up as if to question.

Animal Sounds were transcribed with @o (onomatopoeia) if they were commonplace, recognizable sounds (e.g. moo, woof, neigh, meow). If other other animal sounds were made, they were transcribed using &=imit: "animal" notation.

- With respect to a *MED tier (talking toy): If the animal sound was spoken as a word (e.g. moo) it was transcribed as such. If the toy played a cow sound, it was transcribed as &=imit:cow.