

Acquisition Sketch Project Meeting 6:

Data Processing Part 2 –
Transcription, Translation, Annotation,
Deciphering & Interpreting Child Utterances

April 2024

Agenda

- Deciphering children's utterances
- Transcription
- Glossing children's utterances



Deciphering children's utterances

Deciphering children's utterances

- Children's language is often:
 - Non-target-like
 - Repetitive
 - Very linked to immediate context, the 'here and now'

Deciphering children's utterances

- An unfortunate truth: Some utterances will remain unclear, despite all attempts
- A tolerance for uncertainty is needed!
- Aim for a balance between quality and practicality

Deciphering children's utterances

- To help achieve balance of quality/practicality, decide on the following:
- Number of times to review an utterance (before going with best guess or deeming it unintelligible)
- Conventions for transcribing best guesses
 - Transcribe segment as XXX and put best guess on another tier?
 - Put best guess on main tier but note your uncertainty on another tier?
- Criteria for when to ask for a second opinion

Deciphering children's utterances

- What can help to decipher children's language:
- Engaging someone very familiar with the child and their daily life (e.g. a caregiver)
- *Other children can also be very good helpers, as can the focus children themselves*

Deciphering children's utterances

- What can help to decipher children's language:
- Engaging someone very familiar with the child and their daily life (e.g. a caregiver)
- Repetitions by the child of the same word/utterance
- *Note: repetitions may or may not be contiguous*

Deciphering children's utterances

- What can help to decipher children's language:
- Engaging someone very familiar with the child and their daily life (e.g. a caregiver)
- Repetitions by the child of the same word/utterance
- *Note: repetitions may or may not be contiguous*

Child 3;2

2:17mins



Deciphering children's utterances

- What can help to decipher children's language:
- Engaging someone very familiar with the child and their daily life (e.g. a caregiver)
- Repetitions by the child of the same word/utterance
- *Note: repetitions may or may not be contiguous*

Child 3;2

2:17mins



5:14mins



Deciphering children's utterances

- What can help to decipher children's language:
- Engaging someone very familiar with the child and their daily life (e.g. a caregiver)
- Repetitions by the child of the same word/utterance
- Other participants' utterances before/after child's

2:03mins

Adult: *mamba da **thelput** manawathanu*

'Well I'm going to make a **house**'

Deciphering children's utterances

- What can help to decipher children's language:
- Engaging someone very familiar with the child and their daily life (e.g. a caregiver)
- Repetitions by the child of the same word/utterance
- Other participants' utterances before/after child's
- Nonverbal behavior of child or others present
- e.g. Someone gives orange to child – previous request for food?

Deciphering children's utterances

- What can help to decipher children's language:
- Engaging someone very familiar with the child and their daily life (e.g. a caregiver)
- Repetitions by the child of the same word/utterance
- Other participants' utterances before/after child's
- Nonverbal behavior of child or others present
- Environmental context
- E.g. bird song, a car/plane

Transcription

Preparing to transcribe children's language

- Extra steps to maximize transcription success and speed:
 - 1) Familiarize yourself with the context of each recording
 - 2) Prepare file ahead of transcription
 - 3) Note repetitions

Preparing to transcribe children's language

- 1) Familiarize yourself with the context of each recording
- Watch the recording session before starting to transcribe, to
 - give you an overall sense of the context and what happens
 - help you tune into individual child/children's speech
- If the transcriber wasn't present at the recording session, or the recording was made a while ago, consider playing a short stretch of the video before transcribing

Preparing to transcribe children's language

- 2) Prepare file ahead of transcription
- First pass transcript: leave out interjections, false starts, hesitations, screams etc.
- So transcriber can focus on deciphering children's utterances
- Segment file (or some of file) ahead of transcribing

Figure 2a. Initial segmentation in the Qaqet corpus (for first-pass transcription).

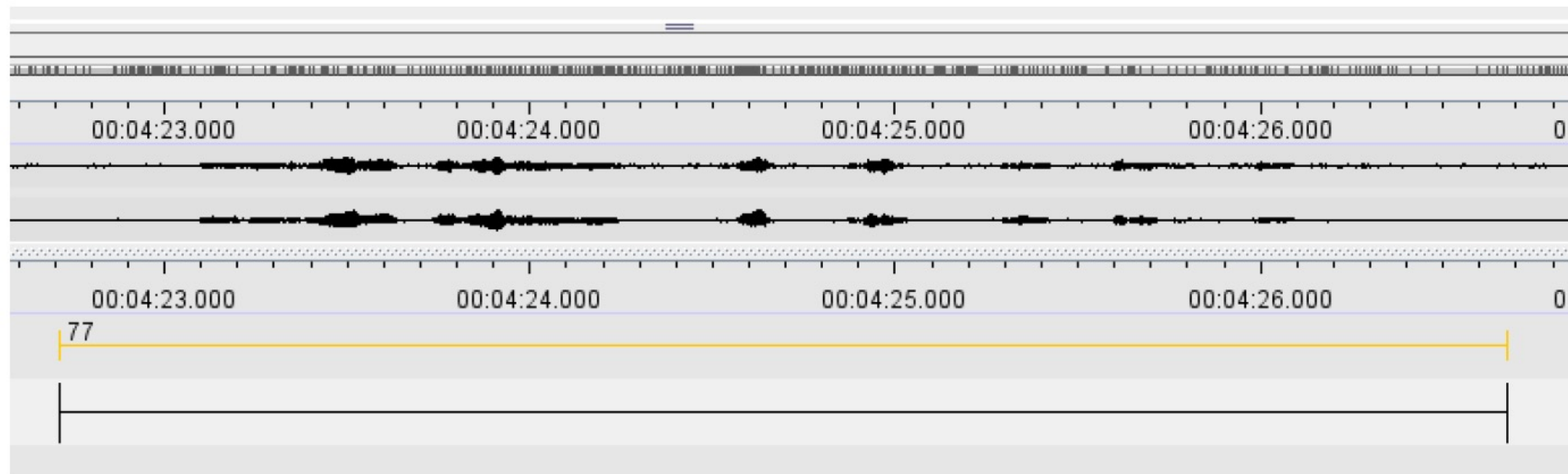
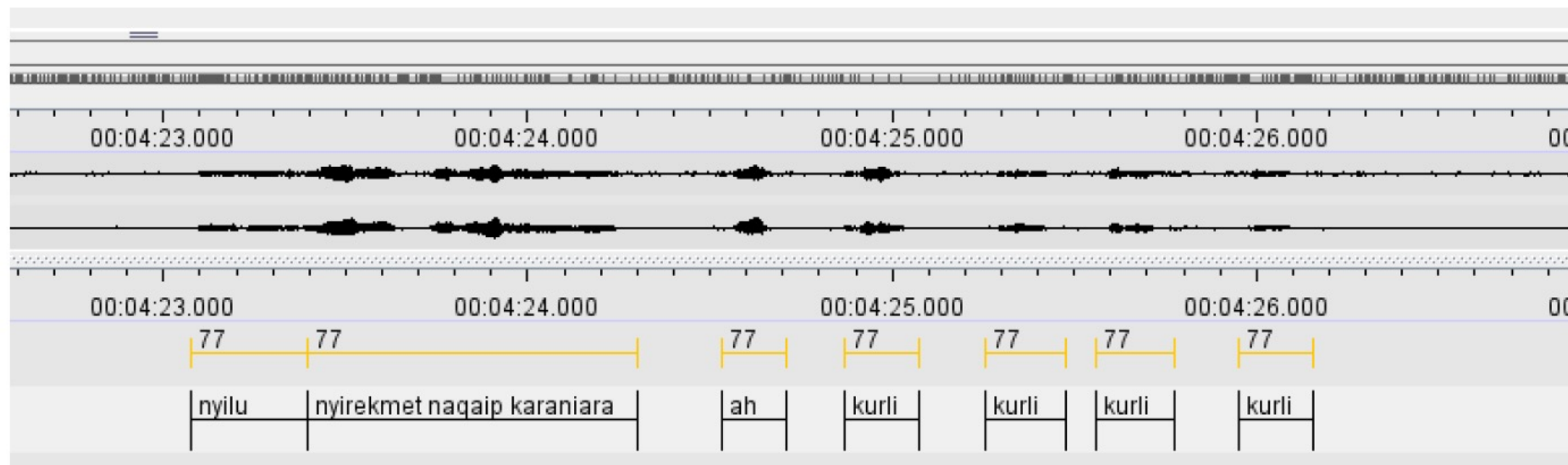


Figure 2b. Final segmentation in the Qaqet corpus (intonation units).



Preparing to transcribe children's language

- 3) Note repetitive utterances
- Use repetitions as a tool when utterance is hard to decipher
- Exclude non-useful repetitions to prevent transcriber fatigue/boredom

Transcribing children's utterances

- Transcription: an iterative process
- The iterative nature of transcription is actually helpful!
- Understanding of individual children's language often improves through the transcription process

Transcribing children's utterances

- Even for first pass transcription, numerous tiers are needed.
- You'll want to record:
 - Child's actual utterance
 - Adult interpretation of utterance
 - Translation
- Probably also:
 - Notes

Transcribing children's utterances

- Child's actual utterance
- Adult interpretation of utterance
- Translation
- Notes

Transcribing children's utterances

- Child's actual utterance
- Try to be as faithful to the child's actual utterance as possible
- Pronunciation: decide how you will represent this in the transcript
- IPA? Standard orthography plus a few extra graphemes? A combination of both?

Transcribing children's utterances

- Child's actual utterance
- Adult interpretation of utterance
- Translation
- Notes

Transcribing children's utterances

- Adult interpretation of utterance
- Transcribers may feel compelled to provide 'proper' version
- Important to allow an outlet for adults to express this
- It could also lead to interesting conversation/comments around common 'mistakes' by children generally
- It is important that transcribers know that it is ok if they are not sure what the child is saying!

Transcribing children's utterances

- Child's actual utterance
 - Adult interpretation of utterance
 - Translation
 - Notes
-
- into national language, English, etc, to make corpus accessible to wider audiences

Transcribing children's utterances

- Child's actual utterance
 - Adult interpretation of utterance
 - Translation
 - Notes
-
- e.g. for generalizable features of children's language ('babytalk')

First pass transcription, Murrinhpatha

The screenshot displays a transcription software interface. At the top, a waveform shows the audio signal over a time range from 00:43:29.000 to 00:43:32.000. Below the waveform, a list of transcription annotations is shown, each with a time range and a label. The annotations are:

Label	Time Range	Text
po@Beverly [466]	00:43:30.500 - 00:43:31.000	* MAMA ngarra ?ngama
tf@Beverly [440]	00:43:30.500 - 00:43:31.000	MUM how (?do i do it)
tn@Beverly [64]	00:43:30.500 - 00:43:31.000	kanhingawu mama ngarra ngama
po@Teresa [834]	00:43:29.000 - 00:43:30.000	tilililman thama
tf@Teresa [816]	00:43:29.000 - 00:43:30.000	say 'spear head grass'
tn@Teresa [52]	00:43:29.000 - 00:43:30.000	
notes [65]	00:43:29.000 - 00:43:30.000	This pron of 'ku tiliman' with extra 'il' used with kids

po= practical orthography; tf = free translation; tn = transcription notes

Transcribing children's utterances

- Second pass/subsequent passes:
- Add everything you excluded at first, e.g. false starts, hesitations, interjections, repetitions. (These are important to include)
- Add extra tiers as needed/desired for annotation e.g. addressee, phonetic transcription,

Final transcript, Qaqet

ref	reference for data file
addr	addressee
trs	transcription
trs-i	interpretation of child utterance
ft	free translation (English)
tp	free translation (Tok Pisin)
tx	text
mb	morphemic breakdown
ge	gloss
ps	part of speech
lg	language
idx	index
nt	notes
sound	link to sound file
dt	date

The screenshot shows the SAM (Speech Analysis Module) interface. At the top, there is a timeline with a red vertical line indicating the current position in the audio file. Below the waveform, a transcript is displayed with various fields corresponding to the legend on the left. The transcript is for a file named 'LongZDL20160112_2_0638' and contains the following information:

- ref@ZDL [718]: LongZDL20160112_2_0638
- addr@ZDL [718]: APA
- trs@ZDL [718]: pupuka
- trs-i@ZDL [718]: apupuqa
- ft@ZDL [718]: grandpa
- tp@ZDL [718]: pupuman
- tx@ZDL [718]: pupuka
- mb@ZDL [835]: pupu -ka
- ge@ZDL [1536]: granny -NC.SG.M
- ps@ZDL [1536]: N -sfx
- lg@ZDL [1536]: TP -Q
- idx@ZDL [1536]: 01344 -00732
- nt@ZDL [718]:
- sound@ZDL [718]: LongZDL20160112_2.wav 897.471
- dt@ZDL [718]: 20/May/2019

Transcription as data collection

- Opportunity to collect more information about an individual child's language, and also ethnographic data
- Can also often spark great conversations about language development more generally, plus local ideas/beliefs/practices regarding language learning
- e.g. discussion of typical 'mistakes' children make at certain ages/stages, discussion of multilingualism in the community



Glossing children's utterances

Glossing child utterances

- How to gloss words that contain partially realized morphemes?
- Tricky, and there is no simple answer.
- 1) Gloss the adult interpretation of the utterances
 - The corpus is more searchable and accessible
- 2) Gloss the child's actual utterance
 - Avoids clouding picture of the child's morphological development
- We suggest option 2), but both choices are valid.

Glossing child utterances

- Example from Qaqet

Glossing child utterances

- Example from Qaqet
- Child attempts to repeat utterance of interlocutor

(1)	child (1;11):	<i>mpapama</i>	target:	<i>gupapaiama</i>
		<i>m=papa-iam=a</i>		<i>gu=papa-iam=a</i>
		DET=papa-DU.M=DIST		1SG.POSS=papa-DU.M=DIST
		'two parents'		'my two parents'
				SAM, p33

Glossing child utterances

- Example from Qaqet
- Child attempts to repeat utterance of interlocutor

(1)	child (1;11):	<i>mpapama</i>	target:	<i>gupapaiama</i>
		<i>m=papa-iam=a</i>		<i>gu=papa-iam=a</i>
		DET=papa-DU.M=DIST		1SG.POSS=papa-DU.M=DIST
		'two parents'		'my two parents'

SAM, p33

- [am] – out of context this would be analyzed as *-am*, but enough info in this context to analyze it as *-iam*
- Initial [m] – very different from target *gu*. More similar to articles *ma* & *ama*. Better analyzed as one of these? If so, which one?? Analytic decision here: *m* glossed as DET

Glossing child utterances

- Example from Qaqet
- No assumptions about what morpheme the child ‘intended’ to use.
- This analytic decision simply makes the corpus searchable (key to detecting patterns)
- This approach also guides against an overinterpretation of ‘errors’
- Impossible to know if something is an ‘error’ from one example
- Such interpretations may obscure patterns in the data

Glossing child utterances

- Analysis of ‘unknown morpheme’ preferable to no analysis
- Example from Murrinhpatha
- Child attempts to repeat utterance of interlocutor

child (3;2)	<i>xxx-aykay</i>	target: <i>thanamkaykay</i>
	<i>xxx-kaykay</i>	<i>thanam-kaykay</i>
	unknown-call_out	2SG.BE(4).NFUT-call_out
	‘XXX call out’	‘you call out’

- Decided on ‘unknown’ despite knowing the target form. Too unclear.

Glossing child utterances

- Record the analytic difficulty and the reasons for your analysis (on a separate tier, e.g. ‘notes’)
- If it is a recurring issue, make a note in a separate document (e.g. in Qaqet: nasal elements in pre-nominal slots) to free up the notes tier.

Glossing child utterances

- Rote-learned forms- a common phenomenon where a child produces a morphologically complex form without having analyzed internal structure
- Likely not enough instances in the sketch to be able to discern whether a form is productive or not.
- This is ok!
- Glossing the form does not assume productivity.



Questions & Discussion