Acquisition Sketch Project Meeting 5:

Data Processing Part 1 – Setup, Transcribers, Segmenting, Tiers

6th March 2024

Agenda:

Selecting a setup

- Selecting and training transcribers
- Segmenting the data

Tiers for transcribing and annotation

Selecting a setup

Two most common options

ELAN plus Toolbox or FieldWorks

- Typically used in language documentation
- Shows sound waves and video on screen
- Tiers can be hierarchically connected
- Best option for annotating visual information e.g. gesture
- https://archive.mpi.nl/tla/elan

CHAT transcription system and CLAN analysis tools in CHILDES

- Typically used in child language acquisition
- Shows sound waves and video on screen
- Tiers not hierarchically connected
- Relatively easy to learn and use
- Dedicated set of analysis tools specific to child language research (CLAN)
- https://childes.talkbank.org

% TEAN 32-78-011-10K/SMRUA/SLC1/MILACTINEAR File Edit Annotation Tier Type Search View Options Window Help Grid Text Subtities Lexicon Come • MAI • Mai	- D X	1	X ELAN 3.7-EX-011_42M_JOINT_C2_MINTACERN.64F	
File Coll Allinolation Her Type Sealch View Opticits Visitation (Her) Grid Test Subtiles Lexicon Come Mail National Sealch View Opticits Visitation (Sealch View Opticits) Mail Sealch View Opticits (Sealch View Opticits) Mail Sealch View Opticits) Mail Sealch View Opticits (Sealch View Opticits) Mail Sealch View Opticits) Mail Sealch View Opticits (Sealch View Opticits) Mail Sealch View Opticits) Mail Sealch View Opticits (Sealch View Opticits) Mail Sealch View Opticits) Mail Sealch View Opticits (Sealch View Opticits) Mail Sealch View Opticits) Mail Sealch View Opticits (Sealch View Opticits) Mail Sealch View Opticits) Mail Sealch View Opticits (Sealch View Opticits) Mail Sealch View Opticits) Mail Sealch View Opticits (Sealch View Opticits) Mail Sealch View	ments Recognizers Metadata Controls		File Edit Apportation Tier Tune Search View Ontions Window Help	
00.00.00.00 Selection: 00.00.01.074 - 00.00.01.379 255 14 10.00.01.00 Selection: 00.00.01.074 - 00.00.01.379 255	Begin TL End Time Duration 000008 000009 000000 000011 000010 000000 000011 000010 000000 000011 000010 000000 000011 000015 000010 000015 000017 000017 000017 000012 000002 000022 000002 000002 000022 000002 000000 000022 000002 000002 000022 000002 000002		Grid Text Subbles Lexion Com • FAI • M Annotation 10 (Figures) 31 (Softman Japanese Hyper) 31 (Sof	ments Recognizers Metadata Co Begin TL End Time Du 00000 0011 00000 0011 00000 0011 00000 0011 00000 0011 00000 0011 00000 0011 00000 0011 00000 0011 0000 0011 0000 0011 0000 0011 0000 0011 0000 0011 0000 0011 0000 0011 0000 0011 0000 0011 0000 00110 0000 0011 0000 0011 0000 00110 0000 0011 0000 0011 0000 00110 0000 0011 0000 0011 0000 00110 0000 0011 0000 0011 0000 00110 0000 0011 0000 0011 0000 00110 0000 0011 0000 0011 0000 00110 0000 0011 0000 0011 0000 00110 0000 0011 0000 0011 0000 00110 0000 0011 0000 0011 0000 00110 0000 0011 0000 0011 0000 00110 0000 0011 0000 0011 0000 00110 0000 001100000000000000000000000000000
				000023000 000022000 00 ++

A sample from the ACLEW project .

oh e square

A sample from the ACLEW project .

Description: With **ELAN** a user can add an unlimited number of textual annotations to audio and/or video recordings. An annotation can be a sentence, word or gloss, a comment, translation or a description of any feature observed in the media. Annotations can be created on multiple layers, called *tiers*. Tiers can be hierarchically interconnected. An annotation can either be time-aligned to the media or it can refer to other existing annotations. The content of annotations consists of Unicode text and annotation documents are stored in an XML format (EAF).

Main features and characteristics of ELAN:

· provides several ways to view the annotations, each view is connected to and synchronized with the media timeline

what the blue cross and th

- · supports creation of multiple tiers and tier hierarchies
- supports Controlled Vocabularies
- · allows linking of up to 4 video files with an annotation document

oh.

- media support
 - builds on existing, native media frameworks, like Windows Media Player, QuickTime or VLC
 - support for audio and video formats depends on operating system, high performance media playback can usually be achieved
- technical

- xds@MA1

- written in the Java programming language
- distributions available for Windows, macOS and Linux
- o open source, the sources are available under a GPL 3 license

				뛽 ELAN 6.0 - LAMP_2013050	02_WF_01.eaf		
	File Edit Anno	otation Tier Type Searc	h View Options Wind	ow Help			
ELAN	CAY	Sanda L			Cria Text S ▼ po@Acacia > Nr Annotation	Subtitles Lexicon Cor Begin Tim	nments Recognizers co e End Time Duration
					117 *** 118 *** 119 * ngarra nhinhi 120 *** 121 *** 122 kura nangkal nuk 123 kura ngay 124 nart ne	00:28:10. 00:28:15. 00:29:41. 00:29:57. 00:30:09. kun * 00:30:19. 00:30:23. 00:30:25.	246 00:28:11.496 00:00:01.250 625 00:28:16.750 00:00:01.125 444 00:29:43.762 00:00:02.318 817 00:29:59.885 00:00:02.068 454 00:30:10.033 00:00:00.579 034 00:30:21.216 00:00:02.182 102 00:30:24.238 00:00:01.136 897 00:30:26.829 00:00:00.932
		00:13:35.622 F4 -14 > >+ >F >1 >	Selection: 00:13:32.79€ ▶I ▶S S → ←	5 - 00:13:34.489 1693 - → ↓ ↑ Selection Mode	e 🗌 Loop Mode Ŵ		ar an 18 f 101 a196 f 111 f = 11 00 − 11 a1 =1000
	LAMP_201 😒	00:13:25.000 00:13:26.000	00:13:27.000 00:13:28.00	0 00:13:29.000 00:13:30.000	00:13:31.000 00:13:32.000	00:13:33.000 00:13:34.000	00:13:35.000 00:13:36.000
	**	· • * * * * *	• ••	••• ••••••••••••••••••••••••••••••••••	· ••••		
	po@Emily	00:13:25.000 00:13:26.000	00:13:27.000 00:13:28.00	00 00:13:29.000 00:13:30.000	00:13:31.000 00:13:32.000	00:13:33.000 00:13:34.000 wakay mananthi warda	00:13:35.000 00:13:36.000
	tf@Emily [440]					that's it, we 've got nothing left!	
	tn@Emily						
	po@Acacia [361]						
	tf@Acacia [332] tn@Acacia						
	[37] po@Tania	mamba nhinhi w	11	dhu-wa kanamkaykay			thangkunu thanamka
	[834] tf@Tania [816]	ok now your turn	+	There you go, he's calling out			say to him 'what are
	tn@Tania						
	notes	-	Dird Call				

Murrinhpatha, data from LAMP project

CHAT

CHILDES	N	Child Language Data Exchange System
CHILDES is the child language compo	nent of the TalkBank system.	
System	Database	Programs
Ground Rules	**Index to Corpora**	CLAN
Contributing New Data	Browsable Database	XML creator and XML Schema
IRB Principles	<u>TalkBankDB database search</u>	Related Software
	Hints on Downloading	
Links	Teaching	Manuals
Other Child Language sites	Topics in Language Acquisition	CHAT Transcription Manual
Research based on CHILDES	Teaching Resources	CLAN Program Manual
<u>Child Language Diaries</u> Data	YouTube Examples	Tutorial Screencasts
	<u>Bibliographies</u>	Overviews, Other Languages
Contact	Phonology and Fonts	Morphsyntax
Brian MacWhinney : <u>homepage</u>	Phon and PhonBank	Universal Dependencies
How to subscribe to <u>Mailing Lists</u>	Unicode and IPA for <u>Mac</u>	MOR manual
	Unicode and IPA for <u>Windows</u>	
Media, CA	Resources	Versions
<u>CA analysis</u>	Building a New Corpus	Derived Corpora and Counts
<u>Digitized video</u>	CCT Computerized Comprehension	XML version of the database
<u>Digitized audio</u>	LEAT Assessment Tool	Database Versioning

File Edit	View Tiers Mode Window Help		100
			-
311	*FAT: i-lla-gi-s ulg ^w mad? •		
312	%mor: 3SM#v lla&PERF~prep gi~pro:obj s&3S n ulg ^w mad?		
313	%pho: il:agis ulg ^w mad ^s		2.1
314	*FAT: managu ad-t t-ẓra inna-m ? •		
315	%mor: pro:int managu comp ad~pro:obj t&3SM 3SF#v zra&PERF	F	
316	pro:poss:kin inna-2SF?		-
317	%pho: manuk at:i t ^s zra in:am		100
318	*CHI: nkki ẓri-ɣ-t . •		
319	%mor: pro:per nkki&1S v zri&PERF-1S~pro:obj t&3SM.		
320	%pho: nk:i zlixt	Paused	02:45
321	%mod: nk:i z ^s riʁt		
322	*CHI: s-all-n-inu i-lla ammas n-yi-nn i-lla-sul wayya ammas n-yi-nr	in 165254	ave
26oct23	3[E CHAT] 311 : Press any KEY or click mouse to stop		
		292874 294140 Repeat 0 msec	3
	and a state sector of the state	a second and a second second by the balance in the second second second second second second second second second	
-Maraket	╫╫╾╂╽╪╓┿┿┾┿┿╫╗┫╔╫╫╫┿┿┿┿┿┿┿┿┿┿┿┿┿┿┿┿┿┿┿┿┿┿┿┿┿┿┿┿┿┿┿	- to	
		r	
			*
s im	a de a contra la colle de la contra de contra de contra de contra de contra de contra de la contra de la contra	many the second way and the second way to be and the second second second second second second second second se	
•	we offer out the sector to some even of the sector is offer a stability of a stability of the sector		*R
-H-I			-V
			_
			A 🗌

Berber, data from Abdellah Elouatiq

Helpful questions to ask

What setup do you already know?

What setup do you have support for (e.g. colleague who can help you)?

What setup is common in your subfield?

Is annotating video important for you (e.g. gesture)?

Is annotating phonetics important for you?

Do you need/want hierarchically connected tiers?

What programs/procedures will you use to analyze the data?

Where do you want to archive the data and what requirements do they have?

Selecting and training transcribers

Selecting transcribers

Typical choices

- You
- Caregivers or family members (alone or with you)
- Other community members (alone or with you)
- Maybe get help from older children and/or target child

Typical criteria

- Fluent speaker of the language
- Able to understand the child's speech
- Knowledge of the child's daily context
- Able to write the language ideally using a computer
- Sufficient time and patience to complete the job (could be divided in parts)
- Acceptable to the family (ethical considerations)

Examples of transcribers

Inuktitut

• Students from the community studying in Montreal, working at researcher's office

Cree

- Mother of one target child who served as interlocutor for all recordings
- IPA transcription students at the university trained in Cree phonology

Mongolian

• Researcher and Master's student in psychology who wanted to gain experience

Sesotho

• Mother of each child working together with the researcher

Chintang

• Community members who had studied linguistics

Supporting transcribers

Provide guidelines

- What to transcribe child speech, caregiver speech, others?
- How to divide intonation units, utterances, etc.
- What to do with repetition, uninterpretable speech, etc.
- Transcribe what child *actually* says, not what you think they intended (that goes on another tier)
- Emphasize need for detail, quality, consistency, etc. as much as is reasonable

Amount of support

- Range from sitting together to fully independent work depends on your situation
- Brainstorm ideas for motivation

Building in "quality control" if reasonable

Why?

- Increase transcriber confidence and knowledge
- Journals may request evidence of transcriber "reliability"

Hold training sessions

- All transcribers transcribe same short portion and compare (e.g. 2 minutes)
- Repeat a few times till cross-transcriber consistency is high
- Discuss and find solutions for differences
- Add solutions to guidelines document

Make plan to calculate transcriber agreement

- Pairs of transcribers blind-transcribe 10% of data
- Calculate Cohen's D or % agreement

Segmenting the data

Select portions to transcribe

Scan each recording to select the portion to transcribe

Continuous chunk of 30 minutes

Maximum amount of child and caregiver speech

Avoid excessive crying, adult-only talk, distracting background noise, etc.

Need to listen to / watch recording – finding dense parts in sound file can be misleading

Prepare selected portion for transcription

Prepare audio (or video) files for your selected setup

Integrate audio (or video) file into the file for transcription

Prepare templates of tiers or have workflow in place to automatically generate them

If the transcriber is computer-savvy and has lots of time:

• Start transcribing, segmenting into units as you go

If the transcriber is less computer-savvy and/or has limited time:

• Segment small portions of speech within your setup so the transcriber can just go to each portion directly and not spend time finding the parts to transcribe

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

00:04:23.000	00:04:24.000	00:04:25.000	00:04:26.000	0
				· · · · ·
	······································		· · · · · · · · · · · · · · · · · · ·	
·····	······	· · · · · · · · · · · · · · · · · · ·	· · · · · · · · · · · · · · · · · · ·	·····
00:04:23.000	00:04:24.000	00:04:25.000	00:04:26.000	00
77				

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

00:04:23.000	00:04:24.000	00:04:25.000	00:0	04:26.000 00
			·····	····
00:04:23.000	00:04:24.000	00:04:25.000	00:(04:26.000 00
77	77		77 77	77
Level .	wirekmet negatin kereniere	ab kurti l	kurti kurti	kurti

Sketch Acquisition Manual, p. 26

Select unit of segmentation to facilitate analysis

Intonation unit (typical in language documentation – Himmelmann 2006b)

- Segment of speech that occurs with a single prosodic contour
- Likely to begin with a brief pause and to end in a clause-final intonation contour
- Not necessarily identical with a syntactic unit, but often corresponds to a clause or a group of closely related clauses
- Can be as small as a single word or as long as a full sentence or more

Utterance (typical in child language – MacWhinney 2021)

- Continuous piece of speech, by one person, before or after which there is silence on the part of the person
- Typically a syntactic unit
- Can be as small as a single word or as long as a full sentence
- See CHAT Manual, chapter 9

Helpful tips

Child language is often repetitive

 Plan how to deal with repetitions – e.g. transcribe as one unit, decide depending on intonation

Child language is often hard to interpret

- Use video context for clues
- Listen to preceding and following utterances for clues
- Plan strategies to reduce frustration e.g. maximum number of listens or amount of time before moving on, symbol/code to mark uninterpretable utterance or utterance to come back to later

If transcription resources are limited

 Think carefully about ways to reduce time and cognitive load – e.g. presegmentation, "removing" repetitions from first-pass transcription, providing motivation in procedure / setting / etc.

Tiers for transcribing and annotation

Potential tiers

Minimum

- Transcription in orthography
- Translation into national language, English, etc.
- Morphemic breakdown, gloss, part of speech

Ideas for optional tiers

- Interpretation of child utterance / likely target utterance
- Errors / child-like productions / differences from likely target
- Addressee
- Notes / comments / situational context
- Phonetic transcription
- Annotation for structures of interest

Helpful tips

Create a tier "template" for each utterance (cut/paste or automatically generated)

First pass = just the orthographic transcription and translation

- this will likely take at least 1 hour per 5 minutes of recording
- maybe also include interpretation of child utterance, addressee, and notes

Gradually add information for other tiers as you need them

Only use the tiers/information you need for your own questions and analyses

Do minimal annotations for the entire five hours with more detailed annotations for parts of the five hours and/or specific phenomena

Different people can do different tiers depending on their expertise

• e.g. mother for transcription, RA for morphemic breakdown and gloss

Murrinhpatha in ELAN + Toolbox



Murrinhpatha in ELAN + Fieldworks

po@Emily			thangkunu *-ayka	y-yu				
tf@Emily			what (?are you) ca	alling out for?			po pi	ractical orthography
[440] tn@Emily			PROMPTED, simil	lar number of s	yllables & m	elody as adult's	tn tr	anscription notes honological notes (child)
ph@Emily			thangkunu yayayk	каууо			addr ad notes no	ddressee (adults)
po@Tania	thangkunu thanamkayka	ay-ya nana	_				* ui	nclear speech
tf@Tania	say to him 'what are you	calling out for?'	_					
tn@Tania	'him' is the bird							
[59]	CDS: Emily							
[233]	They are talking to a bird	(pulupulu) in the tree, Pul	lupulu is also the nic	ckname and tot	em of Emily's	s older brother	1	
notes [67]	[605]	thangkunu		thanam	kay	-ya] I na	-na
	Tania morph-ct-m	why		2SG.BE(4).NFUT	call.out	INTJ	2SG.SAY/DO(8).FUT	3SG.M.IO
	Tania_morph-msa-	INT		CSP	Cov	СТС	CSP	ю
	Tania_morph-txt-m	thangkunu		thanam	kaykay	-уа	na	-na
	Tania_morph-type	stem		stem	stem	suffix	stem	suffix
	Tania_phrase-gls-e	say to him 'what are you calling	out for?'					
	Tania_phrase-segn	232						

Mongolian in CHAT

@Begin

@Languages: mon

@Participants: MOT Mother, CHI Target_Child

- @ID: mon|Sketch-Mongolian|MOT|25;05|female|||Mother|||
- @ID: mon|Sketch-Mongolian|CHI|3;0.|female|||Target_Child|||

@Media: SAM001, audio

- *CHI: Ügüi. •187_5157•
- *CHI: Zovoogoogüi baina. •4075_5296•
- *MOT: Naadkhyg chini khen gedeg yum. •8963_9704•
- *CHI: Melkhii gedeg. •10445_10944•
- *MOT: Melkhii gedeg yum uu? •11416_12184•
- *CHI: Mmkhn melkhii gedeg. •12192_14424•
- *CHI: Yoooyo. •18203_18440•
- *MOT: Öör bambaruush baigaa yum uu? •18461_19603•
- *CHI: Baikhgüi. •19488_20624•

12Sep22[E][CHAT] 32





- ightarrow Audio and Video Time Marks:
 - •4075_5296• time alignment
- → Transcribing directly from an audio file using sonic mode in CLAN
- → Translation, morphemic breakdown, gloss and other tiers will be done later

Mongolian, data from Dorjderem Byambasuren

Berber in CHAT

File Edit	View Tiers Mode Window Help	and the second se			
	x 🖻 💼 🧁 💡	A CONTRACTOR OF THE OWNER	X		
86	*MOT: ffi-iyi-d şşkk ^w ar ffi-iyi şşkk ^w ar ɣ-ɣi . •				
87	%mor: v/ffi&IMP~pro:dat/iyi&1S~dir:prox/d n/sskk"ar				
88	v/ffi&IMP~pro:dat/iyi&1S n/sskkwar prep/y\$adv:loc/yi.	A REAL PROPERTY AND A REAL			
89	%pho: fːijːid sː'kːʷar fːijːid sː'kːʷar вːi		110		
90	*CHI: huh?•				
91	*MOT: ffi şşkk ^w ar yi-d ad asi-y zzit . •			*MOT	mother
92	%mor: v ffi&IMP n sskk ^w ar adv:loc yi~det:dem:prox d comp ad v			*~~~	ala il al
93	n zzit.			*CHI	child
94	%pho: f:i s:`k: ^w ar ʁid ad asiʁ z:it	Paused	26:35	*SIS	sister
95	*CHI: huh?•				
96	*SIS: ffi-t-inn ffi-t-inn . •	1595327	Save	%mor	morphemic
97	%mor: v ffi&IMP~pro:obj t&3SM~dir:dist inn				hreakdown
26oct2	3[E CHAT] 10	0 1 Beneat 0 msec	?		
					and gloss
				%nho	nhonotic
		addabhaithine		700110	phonetic
					transcription
		- 1997 - 11			•
+H			+V		
0		And An and Annua			
5	Lead, et al. attraction of the statistical states and states at the states of the stat	A BAN LI BARAN MANANAN	*		
			-		
_					
Ready		NU	M		

Berber, data from Abdellah Elouatiq

Inuktitut in CHAT

*LIZ:	anisaarta.
%eng:	Let's go out.
%mor:	<pre>VR ani^go_out+VV ADV saaq^quickly+VI ta^IMP_1pS .</pre>
%cod:	\$VER \$ITR:vr
%arg:	<pre>\$YRF \$STR \$INT \$SUB:nul:1:hum:crx:lcn:qun:cxp:ugi:rxx:gsn:prn \$WOR:1 \$MOR:3</pre>
%err:	
%tim:	01:02:07
%snd:	"mae14.sd" 3734737 3736477
%add:	DOM; JIN
%sit:	they try to go out the window
%com:	

%eng	English translation	%tim	time on tape
%mor	morpheme breakdown and gloss	%snd	link to audio sound
%cod	codes for verb types	%add	addressee
%arg	codes for argument structure	%sit	situational context
%err	errors	%com	comments

Inuktitut, Lizzie 2;6.3 Data from Shanley Allen

Cree in Phon extension of CHAT including Praat

Session Editor : B1.01-Daisy-03_08_10*			– 0 ×
le Edit Workspace Session View Media Query Analysis Tools Window Help)		
H C Flay segment	Record: 95	of 886 🚺 🖣 🕨 🕅 Q- Search	
🛓 Media Player	🗕 🛃 🗖 🛃 🔀		- & - <u>-</u>
	₫ 95 Speaker: Daisy	Exclude from searches	
	Orthography	mâu=tâh âhtut=ikiniwi=ch-h	
The second se	Morpheme Meaning	this=loc do=passive.3=0.pl	
	Morpheme Type	p,dem+G.pxl=loc IC.initial=passive=CIN	
- Carton -	Translation	We do it like this.]
L KATALAN -	Segment	003:41.724-003:43.347 > Play segment	
Hostin Hunger	IPA Actual	ondædədɪgɪnʊʧʰ	
	IPA Target	[maʊdatotɪgənotʃʰ	
	Actual Morphology	on=dæ dəd=ɪgɪnʊ=វែា	
	Target Morphology	ˈˈmaʊ=da tot=ɪɡəˈno=tʃʰ	
	Passives Coding	B; nSAP; Act; TI; Conj; do; ihtutim	
with a second seco	Passives Commentary	She is talking about scooping them with the spoon. I don't think it is a peel ûtih=î=kiniwi=ch=h 'to dish or scoop'	er but a spoon. scooping or dishing: â
🕨 🕈 Play segment 🔲 🕈 Export segment 🛛 🔀 Reset 🍳 Zoom in 🔍 Zoom out 📓 🖉 Spectrum 🔤	TextGrid Notes	She is talking about how to peel potatoes. From previous record	
003:42.482 003:41.724	003:42,980	[L
Marine			
No TextGrid found Click here to generate TextGrid tiers from Phon tiers.			
Session Information 🔯 Tier Management 👐 Speech Analysis 🚺 CCLAS Auto Parser	ld: 42fca48c-5a31-4506-ad26-dca	4bdc8fa62	Tier: Passives Coding Group: 1 Character: 35
B1/01-Daisy-03 08 10	4 86 warnings		

Cree,

CCLAS

project

Cree – data from Phon section of Talkbank

CHI: itunt~a~u . 🕨
%pho: t ^h əhesaw
%mod: it∫'daw
%xactmor: t ^h əhes~a~w
%xmodmor: it∫'d~a~w
%xmormea: go~vai.fin~3.sg
<pre>%xmortyp: initial~vai.fin~IIN</pre>
%xtrans: is going

child; orthographic transcription *CHI %pho **IPA transcription - actual** %mod **IPA transcription - target** morphemic breakdown - actual %xactmor %xmodmor morphemic breakdown - target %xmormea gloss of morpheme meaning gloss of morpheme function %xmortyp %xtrans **English translation**

Ani 3;01.18, CCLAS database

https://phon.talkbank.org/access/Other/Cree/CCLAS.html https://sla.talkbank.org/TBB/phon/Other/Cree/CCLAS/Ani

Qaqet in ELAN

- 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



SAM, p. 27

Qaqet in Toolbox

🔁 LongZDL20160112_2.txt	
\ref reference	LongZDL20160112_2 0638
\ELANBegin ELAN Begin	897.471
\ELANEnd ELAN End	898.139
\ELANParticipant Participant	ZDL
\addr addressee	APA
\trs transcript (actual utterance)	pupuka
\trs-i transcript (interpretation)	apupuqa
\ft free translation (English)	grandpa
\tp free translation (Tok Pisin)	<i>pupuman</i>
\tx text	pupuka
\mb morpheme breaks	pupu -ka
\ge gloss (English)	granny -NC.SG.M
\ps part of speech	N -sfx
\lg language	TP -Q
\idx index	01344 -00732
\nt notes \sound sound \dt date	LongZDL20160112_2.wav 897.471 898.139 20/May/2019

SAM, p. 28

Qaqet in CHAT

*ZDL:	pupuka
%int:	apupuqa
%eng:	grandpa
%tkp:	pupuman
%mor:	n pupu=grandparent&TP+nc ka=sg&m&Q
%pos:	n+sfx
%add:	APA
%not:	
%snd:	LongZDL20160112_2.wav 897.471 898.139

%int	interpretation	%pos	part of speech
%eng	English translation	%add	addressee
%tkp	Tok Pisin translation	%not	notes
%mor	morphemic breakdown and gloss	%snd	link to audio file

Questions and Discussion

Good luck in data processing!