

Yvonne Treis (CNRS-LLACAN)

Yoseph Yonas Zecharias (University of Wolkite)

Teshome Dagne Madebo (Durame)

Towards a Kambaata Digital Language Archive

2ND HYBRID MEETING ON CUSHITIC LANGUAGES AND LINGUISTICS, NAPLES, 5-6
DECEMBER 2024

Towards a Kambaata Digital Language Archive

- Preliminaries
- Part 1: Project organization
 - Types of data collected
 - Dissemination
 - Collaborative transcription and translation workflow
- Part 2: A first look at backchanneling devices in the corpus

Report about our collaborative work in progress

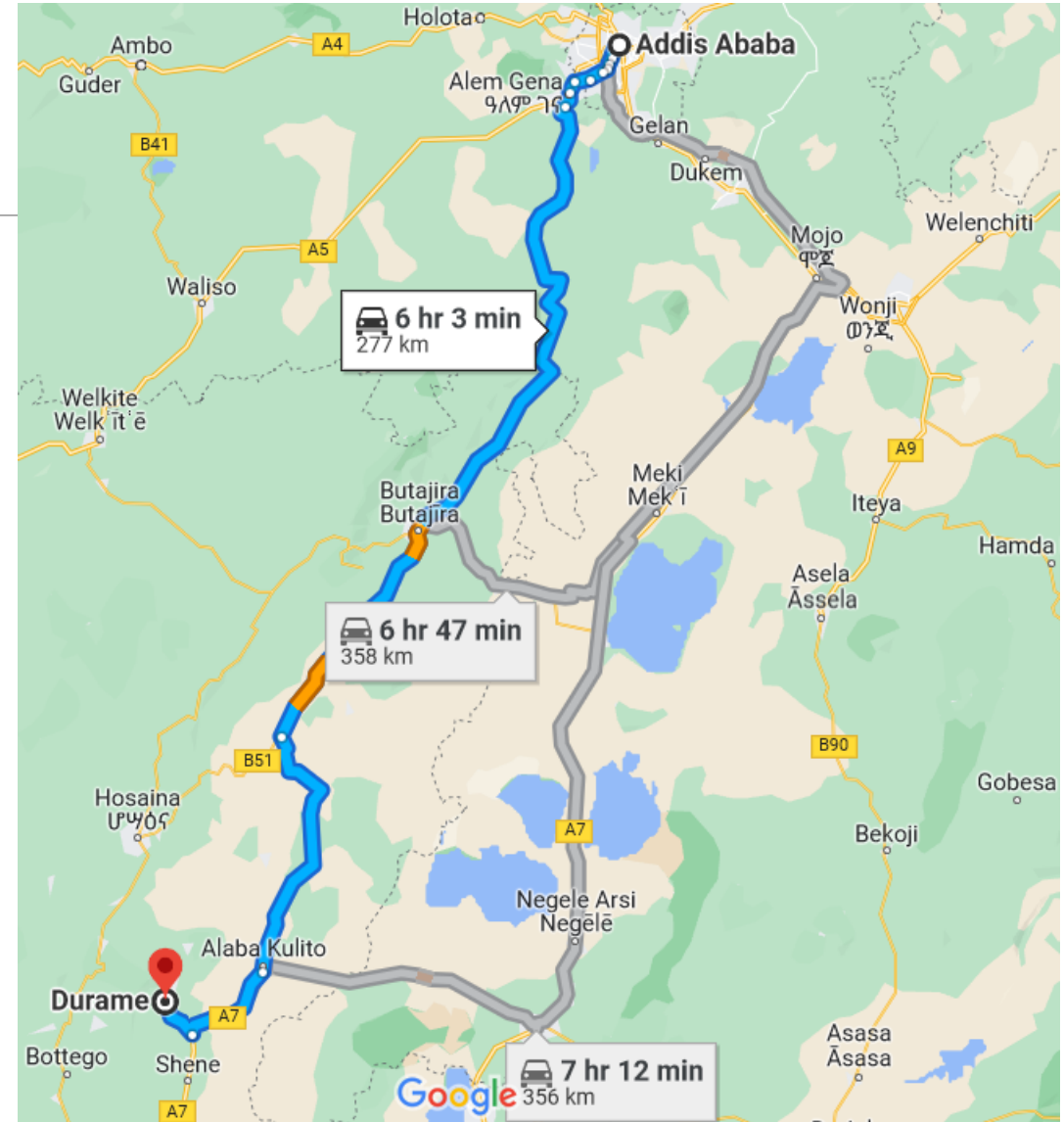


1. Preliminaries

LOCATING THE KAMBAATA AREA, SITUATING THE LANGUAGE

Locating the Kambaata area

- The Kambaata zone (capital: Duuraame) is located in the newly (2023) founded Central Ethiopian Region



Situating the Kambaata language

- Kambaata (*Kambaatissata*) < Highland East < ... < Cushitic < Afro-Asiatic
- Neighbours: Hadiyya and Alaaba (Cushitic) and Wolaitta (Omotic)
- Since the early 2000s, the grammar of Kambaata (cf. my own work) and its lexicon (Alemu Banta 2016) have been studied in detail.
- In 2023, the plan arose to create a **Kambaata Digital Language Archive** with audio-recorded, transcribed, annotated and translated tales, myths, interviews, poems, eulogies, readings etc. (searchable, citeable, freely consultable)



Collaborators of the corpus project

- Language assistants and regular transcribers and translators (= co-presenters)
 - **Yoseph Yonas Zecharias** (University of Wolkite, PhD in Environmental Studies)
 - **Teshome Dagne Madebo** (employee in the office of urban planning)
- Additional help by
 - Deginet Wotango Doyiso (PhD in law)
 - Lidiya Tadesse (Lecturer in Computational Sciences, Wachemo University)
 - Temesgen Senbato (PhD candidate in Linguistics, at Addis Ababa University)
- Numerous native speakers recorded

2. Project* organization

TYPES OF DATA COLLECTED – DISSEMINATION - COLLABORATIVE
TRANSCRIPTION AND TRANSLATION WORKFLOW

*as yet non-financed

Types of data collected



- Goal: Record **conversational** data
- Ideal context would be “coffee conversations”, but impossible to record for an open archive (ethical questions) and very hard to process (too many overlapping voices)
- Compromise in 2023: Conversational interviews
 - Selected main topic 1: Beekeeping
 - Selected main topic 2: Blessings and curses
 - Interviews were conducted by the co-presenters of this paper and recorded at the interviewee’s or the interviewer’s home

Types of data: On beekeeping

- Why conversations about beekeeping?
 - Yoseph Yonas Zecharias is a beekeeper
 - Continuation of a vocabulary collection, written text collection and photo documentation started remotely during the Covid pandemic
 - Potentially of interdisciplinary interest (botany, environmental studies, anthropology ...)
- 6 in-depth interviews (conversational, with a few guiding questions) with practicing Kambaata beekeepers, ca. 360 minutes
- Topics addressed: Beekeeping work cycle, bee life cycle, enemies of the bees, skills and qualities of the beekeeper, equipment and parts of the beehive, bee forage and food, types of honey, climatic changes, changes in the environment and in the harvesting seasons, cultural significance of honey, honey marketing
- Additionally video-taped in 2024: Beehive weaving



Types of data: On blessings and curses

- Why conversations about blessings and curses?
 - Grammatical interest: Kambaata has dedicated “benedictive/maledictive” verb forms
 - High cultural/social significance of blessing and cursing
 - Potentially of interdisciplinary interest (esp. anthropology)
 - Considered very relevant for documentation by members of the speaker community
- 6 in-depth interviews (conversational, with a few guiding questions) with “average” speakers, ca. 420 minutes, by Yoseph Yonas Zecharias and Teshome Dagne Madebo
- Topic addressed: Reasons, contexts, occasions, consequences of blessing and cursing, everyday vs. ritual blessings and curses, alliterating curses, persisting/inherited curses etc.



Archive

- The recordings and their metadata have been archived on the CoCoON platform (<https://cocoon.huma-num.fr>)
- Currently Kambaata 318 “resources” = (parts of) recordings
- Metadata is consultable but recordings under embargo until they are transcribed
- Transcription files are about to be added (test deposits of transcriptions have been made)

Recherche

Mot-clé: kambaata x

Ajouter un mot-clé



Affichage

50 items ▾



1

2

3

4

5

6

7



Interview on blessings and curses (part 2.7)



🔊 audio

"Interview on blessings and curses (part 2.7)" 2023. Kambaata. Teshome Dagne (speaker); Yoseph Yonas Zecharias (speaker); Treis, Yvonne (speaker); Treis, Yvonne (recorder); Treis, Yvonne (depositor); Treis, Yvonne (researcher). Editeur(s): Langage, langues et cultures d'Afrique.

notice



Interview on blessings and curses (part 2.7)



🔊 audio

"Interview on blessings and curses (part 2.7)" 2023. Kambaata. Baqqala Muggooro Kooxo (speaker); Laaqa Muggooro Kooxo (speaker); Teshome Dagne (speaker); Treis, Yvonne (speaker); Treis, Yvonne (recorder); Treis, Yvonne (depositor); Treis, Yvonne (researcher). Editeur(s): Langage, langues et cultures d'Afrique.

notice



Résultats

318 ressources trouvées



Lieux d'enregistrement

Ethiopia

318

Langues

Kambaata

318

Collections

Collections du LLACAN

318

Documentation of Kambaata

318

Documentation of Kambaata blessing and cursing

103

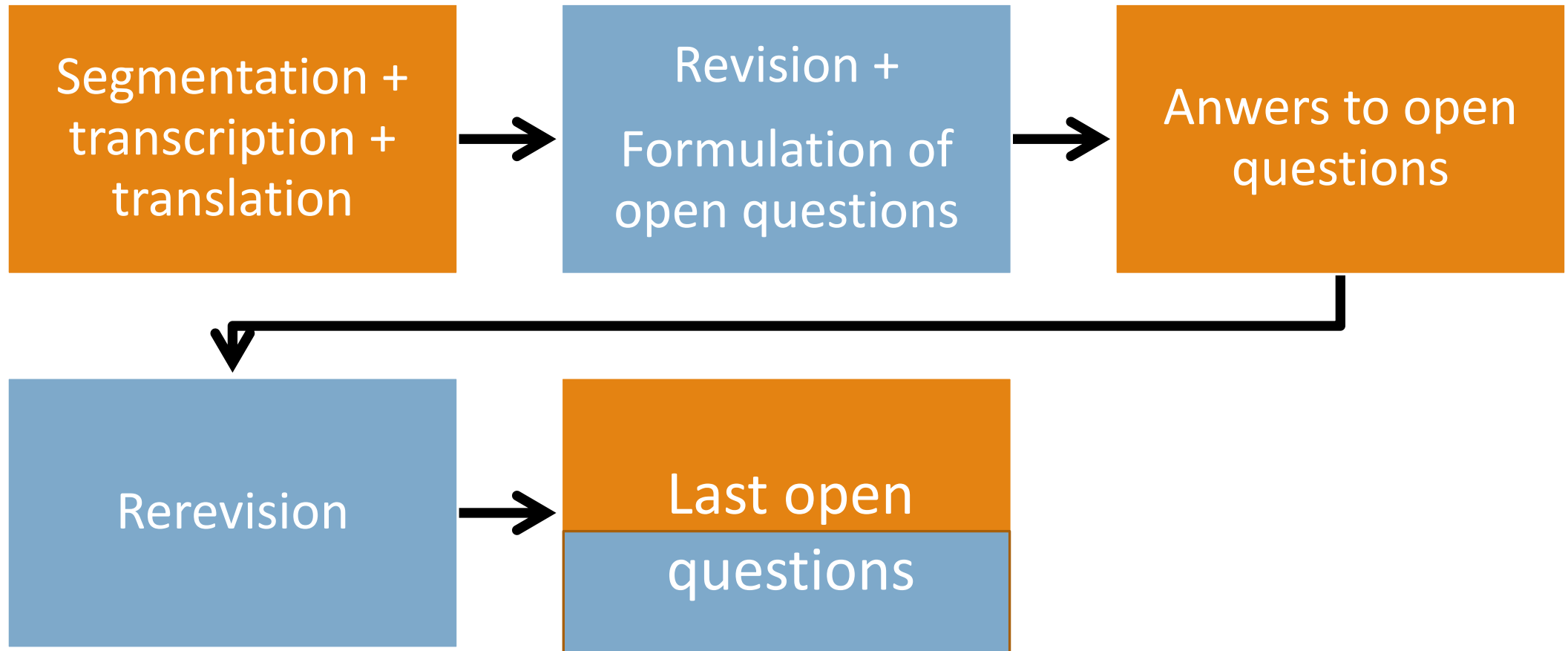
Documentation of Kambaata beekeeping and honey hunting

79

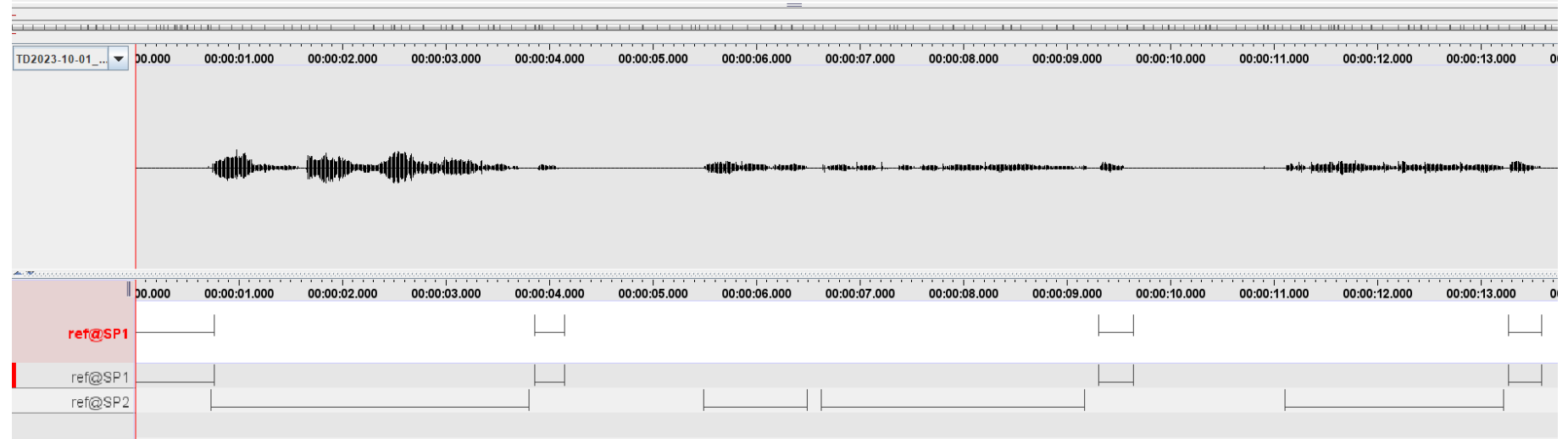
Collaborative transcription and translation workflow

- ELAN training in 2023 in Kambaata
- Preparation before start of remote collaborative transcription
 - Recordings were chunked (max. 5 min) and named acc. to conventions
 - Audio and corresponding empty ELAN-files were left behind in Kambaata
- Principles
 - Transcription team only exchanges .eaf-files (very small size!)
 - Only one person works on one file at a time (version confusion!)

Collaborative transcription and translation workflow



Collaborative transcription and translation workflow



- **Step 1**

- YYZ/TD segment an audio file into intonation units/units between pauses (Annotation mode)
- YYZ/TD transcribe (orthographically) and translate (into English) each segment (Transcription mode)
- Transfer of .eaf-file by email to YT

Collaborative transcription and translation workflow

- **Step 2**

- YT checks and (if necessary) repairs boundaries of segments (Segmentation mode)
- Corrects transcription errors, marks morphological stress, corrects English translation (Transcription mode)
- Adds questions and clarification requests into \nt-tier (= column in Transcription mode)
- Generates unique labels for each segment (= utterance/intonational unit)
- Transfer of (revised) .eaf-file to YYZ or TD

- Step 2a
- Revision

No	Type 3 : tx	Type 2 : ft	tx
32	kánt manní miníiha oróosseent you [rejected?] (the position which is given to me) and you took [what?] to (other) people's home.	@TD: It is not yet clear to me what the mother's complaint is. What does it mean that the son rejected her position? And what did he take to other people's home?
33	atíi hittínta xúudd fá'i_ yéen aguréenno May you also look and remain with bare hands in the same way!" she says.	@TD: Should I write át, átíi or atíi here?
	tx@SP1	ft@SP1	nt@SP1
34	úú .	[BChu]	
	tx@SP2	ft@SP2	nt@SP2
35	tá [mexxita laagat?] . hoolamá(ha) xawaaqeenimbá'a yéemma .	Only this one word. It was said that the mother did not say more.	@TD: Can you check the first phrase. It seems "tí méxxit láagat" to me. Or is it "tá mexxíta laagáta"?
	tx@SP1	ft@SP1	nt@SP1
36			@TD: Can you toggle to Annotation mode to check whether you are able to hear what you say?

- Questions

- Step 2b

- Unique identifier

The screenshot displays an audio analysis software interface. At the top, a timeline shows time markers from 00:03:38.500 to 00:03:41.500. Below the timeline is a waveform representing the audio signal. A vertical red line is positioned at approximately 00:03:38.500. Below the waveform, a list of segments is shown, each with a unique identifier and a corresponding text transcription. The first segment, 'ref@SP1 [110]', is highlighted with an orange box and contains the text 'PM2023-09-21_003_p4_140'. The second segment, 'tx@SP1 [110]', contains the text 'óntu máncu yoodáa kii daafiin uulláta hundánta maassa'aammi_'. The third segment, 'ft@SP1 [110]', contains the text '... even if there are (only) five persons, "Because of you, I bless the whole land", ...'. The fourth segment, 'nt@SP1 [110]', contains the text 'PM here probably refer to verses Genesis 18:26-33.'. Below these are three more segments for 'ref@SP2 [62]', 'ref@SP3 [1]', and their respective 'tx', 'ft', and 'nt' sub-segments, all of which are currently empty.

Segment ID	Text
ref@SP1 [110]	PM2023-09-21_003_p4_140
tx@SP1 [110]	óntu máncu yoodáa kii daafiin uulláta hundánta maassa'aammi_
ft@SP1 [110]	... even if there are (only) five persons, "Because of you, I bless the whole land", ...
nt@SP1 [110]	PM here probably refer to verses Genesis 18:26-33.
ref@SP2 [62]	
tx@SP2 [62]	
ft@SP2 [62]	
nt@SP2 [62]	
ref@SP3 [1]	
tx@SP3 [1]	
ft@SP3 [1]	
nt@SP3 [1]	

Collaborative transcription and translation workflow

- **Step 3**

- YYZ/TD read the questions and type their answers into the \nt-tier in Transcription mode (no changes made in the revised \tx and \ft tiers)
- Transfer of (rerevised) .eaf-file to Yvonne

- Step 3
- Answers to questions

No	Type 3 : tx	Type 2 : ft	tx
39	ûhû .	[BChu]	
40	abbiccámu qoraphphaqqámu yóo'(u) yánnaa ?	(So) it means there is respect of each other and care of each other?	
	tx@SP1	ft@SP1	nt@SP1
41	ââ .	Yes.	
42	abbisáno . qorabáno . abbiseemmasída	It respects. It takes care. If one respects it, ...	@YYZ: Who is the subject of abbisáno and qorabáno, the bees or the beekeeper? @ YT: bees
43	qorabeemmasída	... if one takes care of it, ...	

Collaborative transcription and translation workflow

- **Step 4**

- YT reads YYZ/TD's answers and revises the transcription in the \tx-tier and/or the translation in the \ft-tier, keeps useful notes in the \nt-tier (Transcription mode)

- **Step 5**

- With YYZ: Face-to-face clarification of last open questions on Zoom (once per week)
 - With TD: Clarification of last open question on Telegram (whenever necessary) (same procedure with LT and TS)
 - With DWD: Discussion of remaining tricky things (no fixed schedule)
- **Transcription and translation done** > ready to be archived on CoCoON

Collaborative transcription and translation workflow

- Where are we after 12 months of collaborative transcription?
 - Blessing and cursing interviews (Teshome Dagne Madebo and others):
~ **4h 30 min** time-aligned transcription and translation
 - Beekeeping interviews (Yoseph Yonas Zecharias): ~ **5h 02 min** time-aligned transcription and translation
 - Altogether: ~ **9.5 hours**

3. Case study: Backchanneling

A FIRST LOOK INTO THE ANNOTATED CORPUS

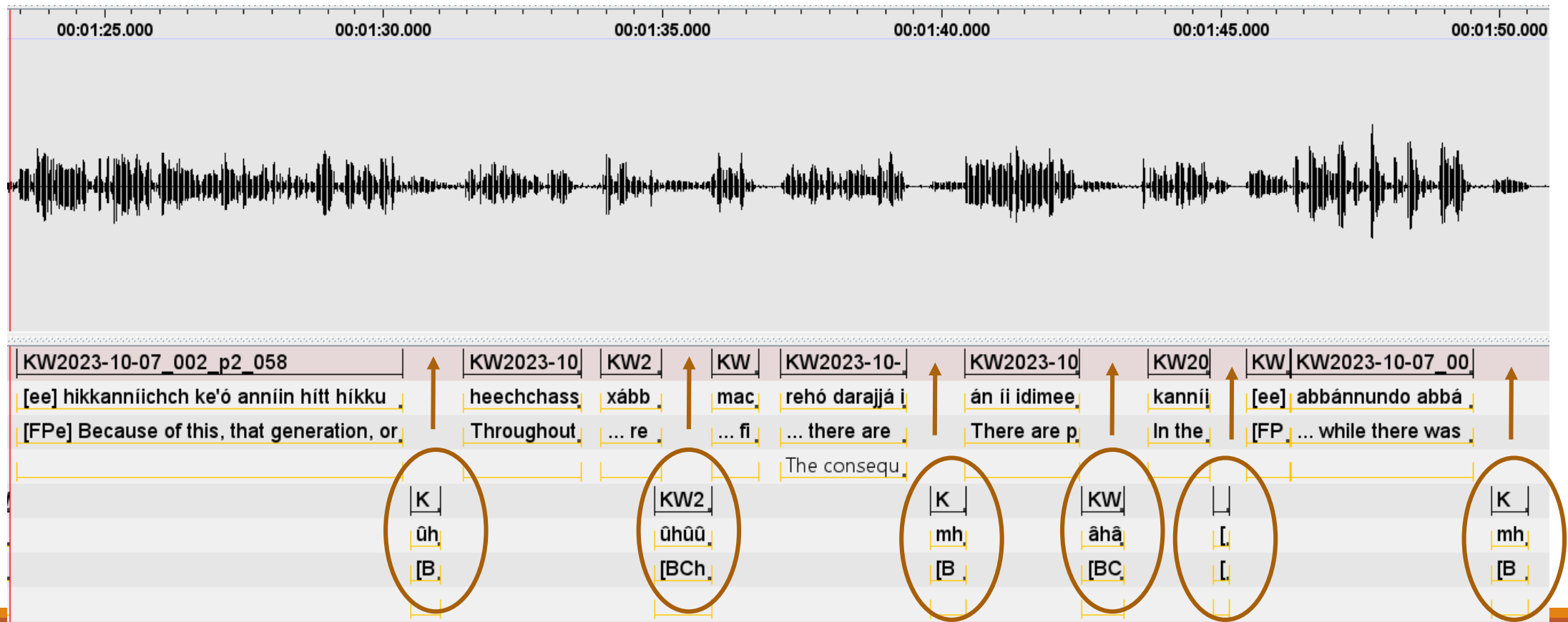
3.1. Introduction

- Backchanneling interjections (a.k.a. response token) establish and maintain communicative contact (Ameka 1992: 114)
- 3962 backchannels have been annotated in the corpus up to today
- Preliminary classification presented below > at a later point, when the transcription of the conversational corpus progresses: quantification possible



Backchanneling: Illustration

- Backchannels are usually perfectly timed to fall into the pauses between utterances





3.2. The default backchannel *mhmm* ~ *ũhũũ*

- The most common backchannel is a nasal or nasalized high back vowel
- Bisyllabic (*mhmm*, *ũhũũ* ...) or monosyllabic (*mm*, *ũũ* ...)
- The vowel quality is central to back and varies in height between more *ũ* and more *ẽ*-like (> phonetic study would be required)
- Default choice

(7) KW2023-10-07_002_p2_037 to 39

SP1		maassa'-áamba'a íi beet-ú bless-1SG.IPFV-NEG 1SG.GEN son-M.ACC 'I will not bless my son.'	
SP2	<i>mhmm</i> BCH[m] 'Mhmm.'		<i>mhmm</i> BCH[m] 'Mhmm.'

3.3. The understanding backchannel *ãhãã*

- The second most common backchannel is a nasalized low back vowel, represented by *ãã* or *ãhãã*.
- In its monosyllabic realization it is not distinguishable from *ãã* ‘yes’ (ex. 8).
- Practical decision w/r/t corpus annotation: The position (after a question or not) is used to distinguish between the backchannel and ‘yes’.

3.3. The understanding backchannel *ǎǎ*

(8) KW2023-10-07_002_p2_112 to 113

SP1	ros-ámm-ee-si-i get_accustomed-PASS-3M.PRF-DEF-NMZ1.M.NOM 'The usual (things)?'	Question	
SP2			ǎǎ yes 'Yes.'



(9) KW2023-10-07_002_p2_014 to 015

SP1	[ee] án kabbóosh woldesambátt-e y-am-am-áamm [FP] 1SG.NOM PN PN-F.VOC say-PASS-PASS-1SG.IPFV 'I am called Kabbabush Woldesambatte.'	Statement	
SP2			ǎǎ BCH[A] 'Mhmm.'





3.3. The understanding backchannel *ãháã*

- The bisyllabic realization *ãháã* seems fairly frequent in contexts in which the speaker signals understanding (cf. its label).
- Typical exchange: Question – Explanation – Backchannel/Understanding

(10) YZ2023-10-14_001_p8_002 to 005

SP1	yaar-áno cry-3M.IPFV 'It cries.'				
SP2		annánn-at distinct-F.NOM	láag-at sound-F.NOM	[yóo'] [COP1.3]	Question
SP1	yáarr cry.3M.PFV.CVB	hir-ch-ú lifesaver.SGV-M.ACC	hir-ch-ú lifesaver.SGV-M.ACC	iill-icc-áno reach-CAUS1.MID-3M.IPFV	
SP2	Explanation	'It cries (to) make a life saver, a life saver come.'			<i>ãháã</i> BCH[a] 'I see.'
					Understanding.



3.4. The narrative backchannel *ee*

- Backchannel *ee* especially found in narrative contexts
- *ee* ‘that’s interesting! now I am curious, continue’
- Very pronounced in the speech of one participant in narrative events (see ex.)
- But also attested in conversations of other speakers

(11) TC2016-02-12_001_p1

SP1	bun-á coffee-M.ACC	gaf-anó-o boil-3M.IPFV.REL-NMZ1.M.NOM	íso-ot 3M.NOM-COP1	
	‘It was him who boiled the coffee.’			
SP2				ee BCH[e] ‘Ee.’
SP1	waas-á enset_food-M.ACC	shol-anó-o bake-3M.IPFV.REL-NMZ1.M.NOM	íso-ot 3M.NOM-COP1	
	‘It was him who baked the <i>waasa</i> -bread.’			
SP2				ee BCH[e] ‘Ee.’
SP1	wo’-á water-M.ACC	eeb-anó-o bring-3M.IPFV.REL-NMZ1.M.NOM	íso-ot 3M.NOM-COP1	
	‘It was him who fetched water.’			
SP2				ee BCH[e] ‘Ee.’

3.5. The ingressive backchannel ↓

- Backchannel expressed as audible inhale (without or with other vocalizations)
- Transcribed as ↓ in the corpus
- **491** attestations in conversations with several different speakers attested in the corpus so far (by 04.12.2024), i.e. $491/3962 = \sim 12\%$ of all backchannels.
- Functional overlap between *mhmm* ~ *ûhûû*, *âhââ* and ingressive backchannel



3.5. The ingressive backchannel ↓

(13) KW2023-10-07_002_p1_039 to 041

SP2	á'nnu 2HON.NOM	kambaat-í Kambaata-M.GEN	roshsh-á=g-íin teachings-F.GEN=SIM-M.ICP
	maass-á=tannée blessing-F.GEN=NMZ2.F.DAT	ros-is-séenunta-nne-et learn-CAUS-2PL.PURP.DS-1SG.OBJ-COP1	

‘It is so that you (HON) teach us about blessings, according to the teachings of Kambaata.’

SP1 ↓
BCH[ingressive]

‘Mhmm.’

SP2	máass-at blessing-F.NOM	kambaat-í Kambaata-M.GEN	roshsh-á=g-íin ... teachings-F.GEN=SIM-M.ICP
-----	----------------------------	-----------------------------	---

‘Blessings, according to the teachings of Kambaata ...’

3.5. The ingressive backchannel ↓

(14) KW2023-10-07_002_p1_103 to 105

SP1 hittig-ú<n>ta ...
like_this-F.ACC<EMP>
'So ...'

SP2 ↓
BCH[ingressive]
'Mhmm.'

SP2 [e] hikku ...
[FP] A_DEM2.M.NOM
'Ee, that ...'



3.5. The ingressive backchannel ↓

- Ingressive backchannels are so far especially reported for Scandinavian languages (e.g. Peters 1981, Pitschmann 1987, Steensig et al. 2022)
- Anecdotal evidence about ingressive backchannels in Ethiopia
 - e.g. Amharic (Ethiopian lingua franca) is included in Eklund's (2015) addenda to his list of languages with pulmonic ingressive speech
- To be investigated: How common are ingressive backchannels in (southwestern) Ethiopia and beyond? Are they an areal feature in East Africa?

Concluding remarks

- Linguistic phenomena will become more easily quantifiable as the transcription processes, e.g. regarding backchanneling
 - Frequency of the different backchannels
 - Preference of different backchannels across speakers
- Our future steps
 - Transcription and translation will be continued
 - Interlinear morpheme glosses are to be added
 - Dissemination on CoCoON (and when glossed on CorpOrAn)
- Our goal: Full verifiability, i.e. our publications should include predominately examples with unique identifiers, directing to an annotated, archived corpus

References

Alemu Banta Atara [Alamu Banta Ataara]. 2016. *Kookaata: Kambaatissa-Amaarsa-Ingiliizissa laaga doonnuta [Kambaata-Amharic-English dictionary]*. Addis Ababa: Berhanena Selam Printing.

Ameka, Felix K. 1992. Interjections: The universal yet neglected part of speech. *Journal of Pragmatics* 18. 101–118.

Eklund, Robert. 2015. Languages with pulmonic ingressive speech: updating and adding to the list. In Malin Svensson Lundmark, Gilbert Ambrazaitis & Joost van de Weijer (eds.), *Working Papers, 55 2015, Proceedings from Fonetik 2015 Lund, June 8–10, 2015*, 31–34.

Peters, Francis Joseph. 1981. *The paralinguistic sympathetic ingressive affirmative in English and the Scandinavian languages*. New York: New York University (PhD dissertation).

Pitschmann, Louis. 1987. The linguistic use of the ingressive airstream in German and the Scandinavian languages. *General Linguistics* 27.

Steensig, Jakob, Auli Hakulinen & Tine Larsen. 2022. Indexing that something is sufficient: Interactional functions of ingressive particles in Finnish and Danish. *Nordic Journal of Linguistics* 45(1). 99–125. <https://doi.org/10.1017/S0332586521000020>.

Thank you for
your attention!

