



CLAN transcription rules for MAIN

MAIN Team at ZAS

21.05.2021

**Leibniz-Centre General Linguistics (ZAS)
Berlin, Germany**

Contact: costmain@leibniz-zas.de

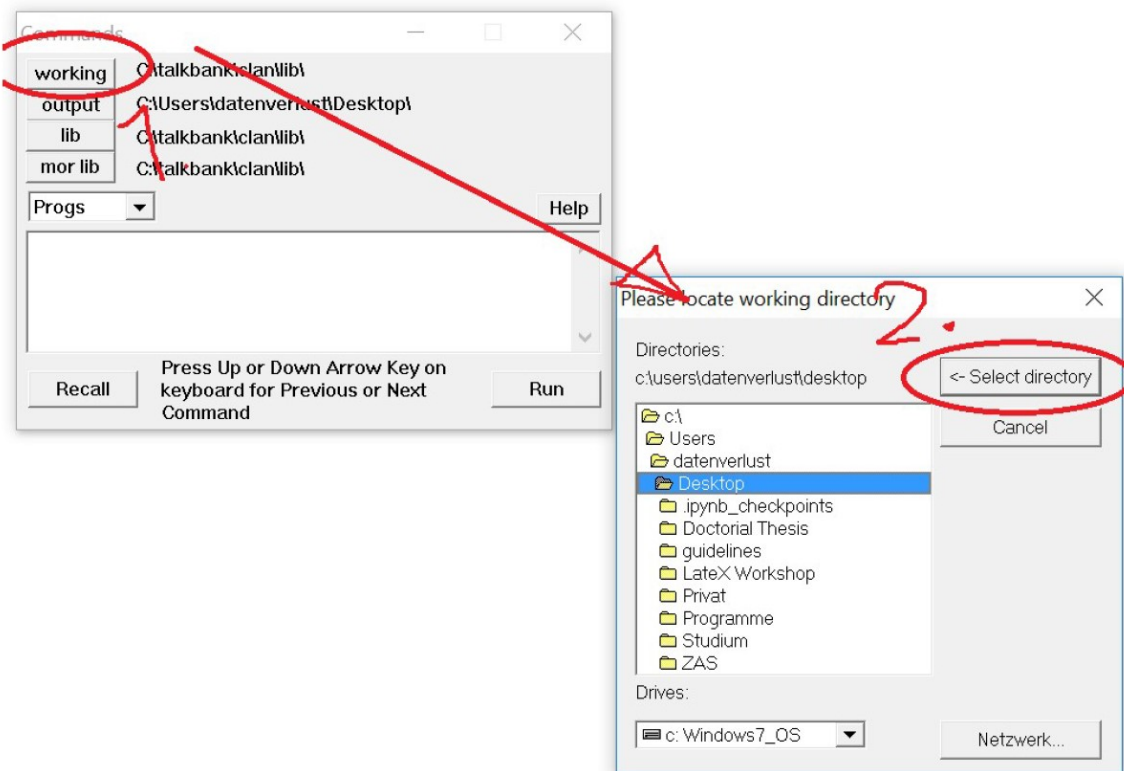
INTRODUCTION TO TRANSCRIBING

Before you start transcribing, you need to download the CHILDES CLAN program. This program allows you to transcribe stories in CLAN format. The CLAN stands for Computerized Language ANalysis.

The present file is structured as follows: first, the structure of the transcripts, the general and detailed instructions for transcribing the audio file and a sample transcript, followed by detailed step-by-step instructions for automatic verification of the transcript and the automatic calculation of the age. At the end you will find examples of transcription decisions.

CHAT transcription

The easiest way is to use the CED (CHILDES Editor) in CLAN for the transcription and save the files with the .cha extension. Download the program at: <http://talkbank.org/clan/> (look for the version corresponding to your operating system - OSx, Linux, Windows).





Before you change anything in the transcript, you should specify the working directory (file path) to CLAN. To do this, open the Commands window with the key combination Ctrl + D. A window with several buttons will open. On the uppermost button says: 'working' here you can specify the file path. That is, in which folder your .cha file is located.

If you have selected the path and passed it with '<- select directory' to CLAN then it is saved for the session (until you close CLAN) and you don't have to use the and you don't have to select the path again later when calculating the age or checking the transcript again.

1. File header

Here is an example of the classic file header and transcript structure. Each transcript starts with a header.

@Begin

@Languages: deu

@Participants: INV Julia Investigator, CHI t052 Target_Child

@ID: deu|BIVEM post3|INV||||Investigator|||

@ID: deu|BIVEM post3|CHI|6;4.23|male|||Target_Child|||

@Birth of CHI: 03 - FEB - 2009

@L1 of CHI: Turkish - German bilingual

@Media: t052_COST, audio

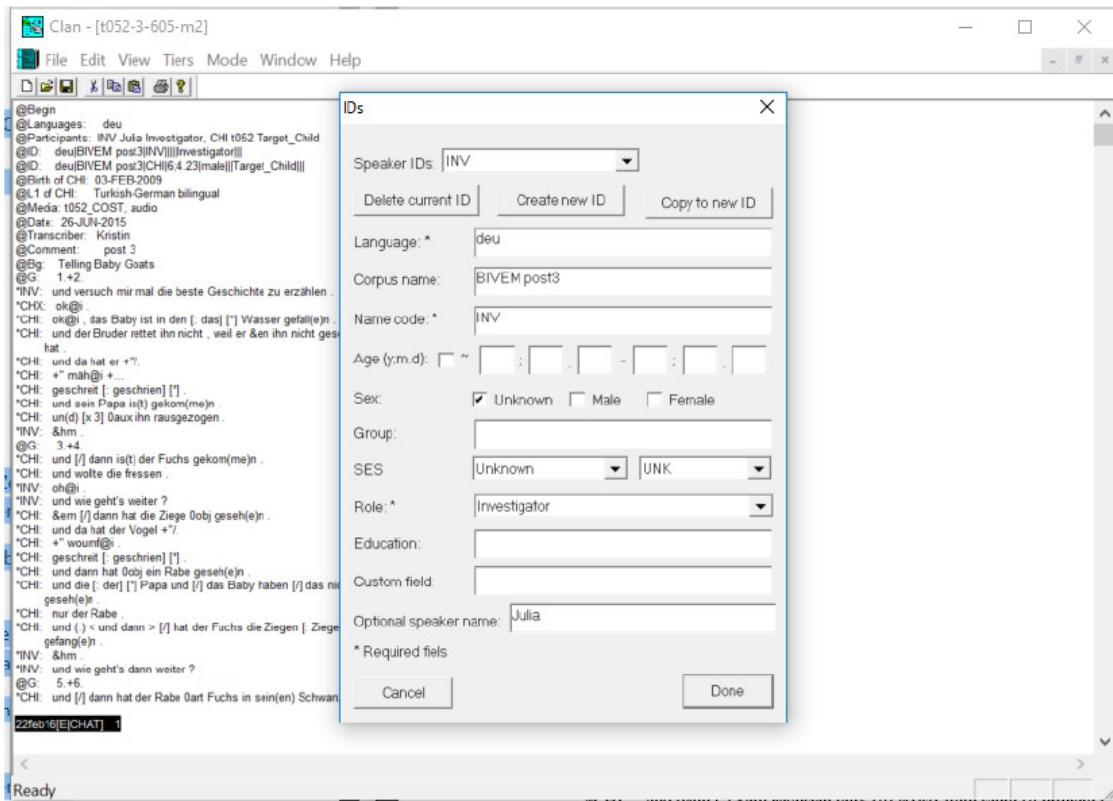
@Date: 26 - JUN - 2015

@Transcriber: Kristin

@Comment: post 3

@Bg: Telling Baby Goats

The header is edited via the menu item 'Tiers' -> 'ID header' in the header. The lines language, Participants and ID should not be changed manually!



Enter the information for LANGUAGE, CORPUS NAME AND NAME CODE.

For the tab about AGE enter as, 6;4.23 → it is the calculated age of the child as of the test day. Here, 6 stands for years, 4 for months, and 23 for days.

- If you want to write the date in the header, the entry should look like this 03-FEB-2009 → important here is that the month always consists of the first three letters of the English month. This can be a test date or birthday date of the test person

@Birth of CHI: 03 - FEB - 2009

@Date: 26 - JUN - 2015

Then follows the story, which, depending on the turn, starts with *INV: , *CHI: , or *CHX: . In addition, the story is divided according to the images in each transcript. See sample transcript at the end of the file.

2. General instructions for CHAT transcription

- The children's stories from the video/audio file are transcribed as accurately as possible. We start with the last statement of the interviewer. (Often this is: 'Now tell me the most beautiful story from the pictures' or similar).
- Phonological, morphological and lexical errors of the child are specially marked (see later in the section-Advanced notes). Colloquial utterances do not have to be improved.
- Each file starts with @Begin and ends with @End.
- After @Languages:, *INV:, %com: etc. the TAB key must be used (not the space bar!).
- On the main/speaker tier, the child is indicated by *CHI: (or *CHX:) and the investigator by *INV:.
- At the end of each line there is a punctuation mark after a space: , . ! ?
- Each complete sentence is written on an extra line. Subordinated Sentences are written in the same line. However, coordination sentences are written in different lines.
*CHI: The cat saw the birds and wanted to eat them .
*CHI: The birds did not know that a cat comes .
*CHI: The cat saw the birds .
*CHI: and he wanted to eat them ,The child's words and sentences should be transcribed as accurately as possible.
- After completion of the story transcript, the transcript should be named according to the following scheme: Child's abbreviation, MAIN, testing point (if several), age rounded. In addition, you can add the story (bb,bg,c,d) or other information relevant to the study.

For example: r100-MAIN-2 (testing point)-406 (406 stands for 4 years and 6 months).

3. Advanced notes

- Sound imitations are marked as onomatopoeias: @o, e.g.: wuffwuff@o// tschiep@o
- exclamations/shouts: **aah@i**
- Interjections (if syntactically relevant): **pst@i, stop@i**; also oh, ach, ohje, oops are marked with @i, so **oh@i, ach@i** etc.; oh my god is a borderline case, possible would be oh@i my God; otherwise marked as not relevant by &oh &my &gott.
- ok should be written consistently **ok@i**
- filled pauses (deliberative or hesitant hm, um, etc.) are written uniformly to: &em
- no differentiation should be made between affirmative and negative **&hm** -> both &hm
 *CHI: and then the mama bird comes flying back .
 *INV: &hm, super !
- Verification questions: or, right, no, not are written down as they are produced
- Pauses: 21,22,23,... (...) We have decided that we will really only transcribe LONG and filled pauses. Say: (.) are long pauses
- unintelligible word: @x or XXX if not understood acoustically
- Omitted words: are completed by marking part of speech or clause. Osub, Overb
- For @G: 1st + 2nd etc. no spaces before and after the plus sign.
- %com: for comments within the text. e.g.: When the child laughs, etc.
- Error: [*]
 CHI: The kat [: cat][] wants to climp [: climb][*] the tree .
- Different word forms are indicated by [: correct word][*] (as a marker for an error).
 CHI: he fallet [: falls][] down .
- Errors are corrected during transcription [: correction][*] & self-correction by child
 [//]:
 *CHI: The kat [//] cat wants to climp [//] climb the tree .
- < > is needed if the correction refers to more than one word
 CHI: The kat [: cat][] <wont to climp> [//] wants to climp [: climb][*] the tree .
 That is, behind what is corrected and in front of the correct word.
- Rephrasing: [///]

- *CHI: The kat [: cat][*] <starts to climp> [///] wants to climp [: climb][*] the tree .
- **tex(t)** Marking a missing element in a word.
*CHI: There is a wo(l)f .
 - Overlapping of several utterances: +<
*CHI: There is a fox .
*INV: +< What is there ?
 - Aborted utterances:
 - +... Incomplete but not aborted utterances.
 - +/. An utterance aborted by another speaker.
 - ++ Completion of another speaker's aborted utterance.
*CHI: look here, here are +/.
*INV: what can you see there ?
*INV: ++ oh@i, birds are here .
 - +//. Own aborted utterance
 - +, completion of a self-aborted utterance
*CHI: oh@i I know &au +//.
*CHI: +, oh@i here !
 - +"/. Direct speech from next line
 - +'' Start of direct speech
*CHI: and then the birds say +"/.
*CHI: +'' chirp@o , chirp@o
 - Repetitions: [/] for one repetition, [x 3] for 2, [x 4] for 3, etc. - with space between x and the number.
 - Arrangement for coding codeswitching:
 - o German word in English: Vogel@s:deu
 - o Russian word in English: slovo@s:rus
 - o Turkish word in English: turwort@s:tur
 - o German word with Russian inflection: fuchsu@s:deu+rus



4. Final check

- @Bg and @Eg must match, that is, after the colon must be the same: Telling Baby Birds or Telling Baby Goats, Telling Cat or Telling Dog
- After you have finished transcribing, you must check if CLAN can process your file and if there are no more errors. To do this, run the CLAN check. **THIS IS VERY IMPORTANT!** To do this, press **esc+L**
- In the Command window you can also calculate the age of the children automatically.
Input: `dates +b00-xxx-0000 +d00-xxx-0000 filename.cha`
where +bxxx stands for @Birthdate and +dxxx for @Date
This opens a new file (.cex) in the same folder, which contains the date of birth. The date should be copied out and pasted into the .cha file. Please always save **ONLY** .cha files.