



# Psychology Resources



## GETTING STARTED WITH CLAN

### CONTENTS

<b><i>Getting started with CLAN</i></b> .....	<b>1</b>
<b>What is CLAN?</b> .....	<b>2</b>
<b>Features</b> .....	<b>2</b>
<b>download</b> .....	<b>2</b>
<b>Prepare your files</b> .....	<b>2</b>
<b>Open CLAN</b> .....	<b>3</b>
<b>Header Lines</b> .....	<b>3</b>
<b>Open Video</b> .....	<b>4</b>
<b>Open Waveform Audio</b> .....	<b>4</b>
<b>Start transcribing</b> .....	<b>5</b>
<b>New lines</b> .....	<b>5</b>
<b>Special characters</b> .....	<b>5</b>
<b>Overlap</b> .....	<b>5</b>
<b>Dependent Tiers</b> .....	<b>6</b>
<b>Saving</b> .....	<b>6</b>
<b>Common Problems</b> .....	<b>6</b>

## WHAT IS CLAN?

CLAN (Computerized Language ANalysis) is a cross-platform program designed by Brian MacWhinney and written by Leonid Spektor for the purpose of creating and analysing transcripts in the Child Language Exchange System (CHILDES) database. CLAN is open source software and can be freely downloaded.

## FEATURES

The CLAN programs include facilities in five different domains:

1. CLAN includes an editor that focuses on the creation of links between words and utterances in the transcript and segment of the related audio or video media. CLAN provides four methods to facilitate this process. The SoundWalker facility emulates the back and forth actions of a transcribers foot pedal, but using keystrokes. Sonic CHAT provides careful segment link from a waveform. Transcriber mode uses the pressing of a space bar after the completion of an utterance to facilitate transcription, and finally time marks can be edited and produced by hand entry.
2. CLAN provides all the basic tools of corpus analysis such as key-word and line, concordance, frequency counting, partial regular expression search, and so on.
3. CLAN provides additional analysis programs for between-speak contingency patterns, utterance and word length, cooccurrence clusters, and so on.
4. For Qualitative Data Analysis CLAN has program such as GEM for marking special segments of text, Coder's Editor for applying a coding system, CA format for Jeffersonian CA transcription.
5. To support the use of TalkBank data in the clinical setting, CLAN includes programs like EVAL and KIDEVAL that compare individual subject and groups with a large comparison database in one or more of the TalkBank corpora databases.

## DOWNLOAD

<https://dali.talkbank.org/clan/>

Quick Start Guide

## PREPARE YOUR FILES

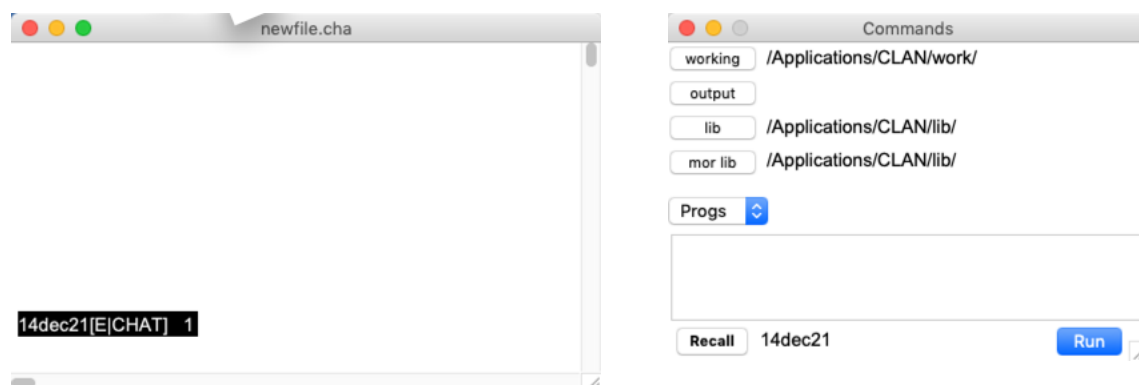
Create a folder to contain your file to be transcribed (this will become your working directory). Your transcript must also be saved to this directory

If you are working on your own laptop/Mac this could be on your desktop, if you are working on a PC on campus it is best to save to your I:drive

## OPEN CLAN

When you open CLAN it may open up in one of two ways.

Directly with a new file ready to transcribe, or a Command window



Using the **Command window** you can set the **working directory** by clicking the working button and navigating to the folder that you have created in the first step.

Next you should set where you would like the transcript output to be saved by clicking the output button and navigating to the folder you created. Keeping files organised is vitally important for your projects

Now click on the main CLAN menu at the top of the screen **File→New** to open a new file

It is a good idea to save regularly, and now is a good time to make the first save **File→Save As**. If you have used Command to set the working and output directory, it should automatically default to the folder that you have specified. Remember that your media file **MUST** be in the same folder as the transcript for you to be able to transcribe it

## HEADER LINES

There are four types of header lines that you will use

@Begin	Do not add anything here
@Options: CA	(@Options→Tab key→CA for conversation analysis)
@Media:	@Media:Tab key→[Insert media file] File→Open→select the media file you wish to transcribe. OR type in the name of the media file after pressing the Tab key (the name of the video/audio file without the file extension, followed by a comma, a space and then either 'video' or 'audio')
@End	Do not add anything here. The transcription should be done <b>above</b> this line. It is always the last line of the file.

Next you will need to add speakers using **Tiers**.

Go to **Tiers** → **ID headers** and fill in the fields marked by an asterisk. The data you input will be used to generate values for the following headers: @Languages, @Participants, and @ID, which are also mandatory header lines and will be generated automatically.

*Note that languages should be given in three letter codes (e.g. 'eng' for 'English'). You can add all participants right away (using the Create new ID option) or click Done and return and do more participants later*

Run the (esc+L) check command to ensure you have filled everything out correctly and save the file

#### OPEN VIDEO

Select **Mode** → **Transcribe audio or video** (or press F5). The audio/video file opens and plays. Click anywhere in the transcript editor window to stop playback

#### OPEN WAVEFORM AUDIO

Select **Mode** → **Sonic Mode** or press Esc+0. The audio of the data file (whether video or audio) will be displayed as a graph ('waveform') on the screen. When using a video file, the first time you select Sonic Mode, you will see CLAN exports the audio from the video file (exporting video). Wait for the program to finish this process before you move on.


Adjust the amount of audio visible on the screen with +H and -H (horizontal)

Adjust the visual amplitude of the audio with +V and -V (vertical)

Should the audio be stereo, and you would like to combine both channels (default), select **Edit** → **CLAN options** → **Show mixed stereo sound wave**

## START TRANSCRIBING

Create a line before '@End'

Highlight a section of the audio by using the mouse. Listen to it again by selecting  and clicking on the highlighted area

Adjust the length of the highlighted section by selecting Shift and clicking to the left or right of the highlighted section. Clicking outside the original selection expands the selection, clicking inside the selection shortens it

Link the highlighted section to the transcript: you either click on the button to the left of the waveform display (·S), or use the shortcut cmd+i (Insert bullet), to add a bullet point (ctrl+i for Windows).

Generate a speaker ID in the left margin, either manually or (better) with the following shortcuts: cmd+1, cmd+2, cmd+3, etc. (ctrl+1 etc. for Windows)

Transcribe the audio

Once a bullet has been added to the transcript, you can 'cmd + click' on it to play the linked video/audio section.

Triple-click a bullet to highlight the corresponding section of the waveform display. This is also what you do when you resume transcription after a break, where you have saved, closed, and reopened the file

## NEW LINES

A new speaker requires a new line. Hit 'enter' to start a new line. Generate a new speaker ID in the same way as above. Pauses between speakers should always have their own line

## SPECIAL CHARACTERS

Occasionally you may need to use certain characters that have been designated to represent certain aspects of the data (e.g. overlap, cut off) or of the transcription process (e.g. inability to hear what is said). The Special Characters menu can be opened in **Windows** → **Special Characters**. To add a special character, simply double click on it

## OVERLAP

Overlapping speech is marked by the special characters provided for this. It is no problem that bullets overlap because of two or more speakers speaking simultaneously. You can 'open' bullets by typing Esc+A, which will show the beginning and end of a bullet in milliseconds from the start of the recording. You can fine-tune bullets by editing them manually

## DEPENDENT TIERS

Between the transcription lines, you may insert comments or descriptions of embodied actions by adding Dependent Tiers. These are generated as %com: or %ges: in the left margin, or some other three letter code following a %-sign

## SAVING

It is advisable to set CLAN to periodically auto-save your work. Select Edit > CLAN Options. In the top box, set the checkpoint to, for example 100. This will automatically save your work after every 100 characters

## COMMON PROBLEMS

Below are some things you need to know to make working with CLAN a lot easier.

1. No spaces or periods “.” in media file names. If you have spaces or special characters in your media file name, CLAN can't open the video or audio
  - a. Resolve this by replacing spaces and periods (or any other special characters) in the file name with '-' or '\_'. Make sure to replace them in the @Media header as well as the actual file name
  - b. Only use letters and numbers in your file name.
2. CLAN crashes unexpectedly (sometimes). Sometimes CLAN crashes unexpectedly. Unfortunately, CLAN doesn't automatically save before it crashes, therefore you should enable the **checkpoint function** that saves your transcript after a set amount of characters written
  - a. To enable checkpoint you have to: Go to “Edit” then “CLAN Options”
  - b. In the window at the top is a field. Enter a number – that number is the amount of characters you have to write before CLAN auto saves your transcript
  - c. It's a good idea to still save your transcript on a regular basis the old fashioned way (ctrl+s // cmd+s)
3. CLAN sometimes doesn't export audio properly. If you are working with a video and you want to use Sonic Mode then CLAN automatically exports a .wav audio of the video. Sometimes that .wav file is corrupted and doesn't work correctly.
  - a. Resolve this by deleting the .wav audio file from the folder and open Sonic Mode again (= export the audio from the video again)
4. The transcript and media file have to be in the same folder
  - a. As it says – the transcript and the media file(s) have to be in the same folder, otherwise CLAN can't play the audio / video file