

LANGSOC LAB
Roskilde University

CLAN Manual

- for workshop training

The screenshot displays the CLAN software interface. On the left, a transcript of a conversation is shown with line numbers 25 to 57. The transcript includes speaker labels (*FRA: and *ROS:) and phonetic annotations in brackets. A yellow highlight is present on line 27. Below the transcript is an audio waveform with a yellow background. On the right, a video window shows two people sitting on a stage in front of a screen. Below the video, a list of phonetic annotations is displayed, including pitch and intonation changes like 'shift to high pitch; F1 up-arrow' and 'falling to low; F1 5'. A yellow highlight is also present on one of the annotation lines.

Spencer Hazel
Janus Mortensen

INTRODUCTION

This document includes a number of introductory instruction exercises that provide users of the transcription linking software CLAN entry points into using the tool.

The various sections included here were produced as materials for workshops led by Janus Mortensen and Spencer Hazel, which sought to provide researchers with the skills to use the tool to benefit their research activities.

In addition, the instructions included here also include a number of sections drawn up for particular research projects with which we have been involved as consultants over the years. We include them here as they may also be of interest to others designing new research projects, who may like to know about potential analytic tools available in CLAN with which the data can be treated.

We do not claim that what is included here is a comprehensive overview of CLAN, but rather ways in which we and our colleagues have been using the software for our own work. For a more thorough description of the tool, we refer users to the manual produced by Brian MacWhinney, available at <http://chilides.psy.cmu.edu/manuals/CLAN.pdf>

Exercises related to the materials included here can be found in the accompanying workshop folder. The exercise data included for this was sourced from Youtube, and made available through a Creative Commons license. The original clips can be found here: <https://www.youtube.com/watch?v=JhrumREAjA>

Janus Mortensen
Spencer Hazel

This manual was produced in conjunction with the CALPIU CLAN transcription manual (2009), written by Anne Fabricius, Hartmut Haberland, Spencer Hazel, Julie de Molade, Janus Mortensen and Bent Preisler. Further guidance and support came from Johannes Wagner at the University of Southern Denmark. Further thanks goes to Brian MacWhinney and Leonid Spektor at Carnegie Mellon University, who have always been helpful and supportive in responding to queries, and in developing the software further in response to suggestions by other users in the community.

NOTE: Some commands differ slightly between Mac OS platforms and Windows. This usually relates to the use of cmd (Mac) and ctrl (Windows). CLAN does work on both.

NOTE: It is advisable to get frequent updates of CLAN from <http://chilides.psy.cmu.edu/clang/> and perform frequent checks of you file

TABLE OF CONTENTS



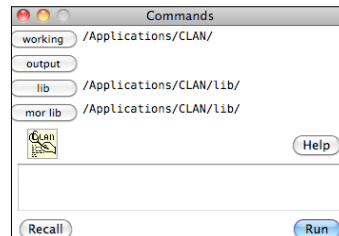
Getting Started in 10 steps	4
Measuring pause lengths	6
Colouring keywords	7
Indent	8
Making collections: GEMS	9
Counting GEMs	10
Hiding and displaying tiers	11
Piping	12
Timedur(ation)	14
Embedding still images	15
Thumbnails: working with gesture annotation	16
Searching and coding with COMBO	20
Exporting to ELAN	23
ELAN- adding subtitles to movie files	24
Appendices:	25
- List of a selection of special characters for transcription	
- Function key commands for transcription	
- Optional Header lines	



CLAN: Getting Started in 10 steps



1. Place the video (.mov) or audio (.wav) file to be transcribed in a folder.
2. **Open** CLAN.
 - a. A command window will open > close it once you have checked the library is set to CLAN/lib.
 - b. Open a new file (File > New)
 - c. **Save** the new CLAN file to the folder where you have placed the video/audio file, giving the CLAN file the same name as the video/audio file when you do so (without the .mov/.wav extension).



3. **Header lines:** you must now fill out the Header Lines for the CLAN file.

A **Header Template** is available, from which a selection of useful headers can be copied.

- a. Open the file 'Header Template.cha'
- b. Select all (Edit > Select ALL), **Copy**, and **Paste** into the CLAN file you are working on
- c. Close Header Template
- d. Fill out the required information (Note: except for Begin and End, all others should be followed by a tab, then the required information).

```
1 @Begin
2 @Options:
3 @Media:
4 @End
```

Mandatory steps

- i. **@Begin** Do not add anything
 - ii. **@Options:** CA
 - iii. **@Media:** The name of the video/audio file (without extension).
This is followed by a comma, a space and then either 'video' or 'audio'
 - iv. 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.
 - v. **@End** Nothing follows this. The transcription should be done above this line. It is always the last line of the file.
- e. Run the (**esc+L**) check command to ensure you have filled everything out correctly.
 - f. **Save** the file
4. 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.
 5. 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.

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


6. Start transcribing.

- a. Create a line before '@End'

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



- c. 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.



- d. 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).

- e. 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)

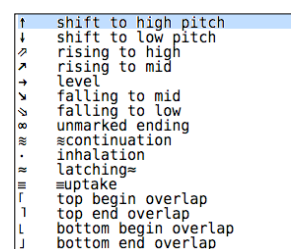
- f. Transcribe the audio

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

- h. **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.

7. New lines. Speaker change should always occasion 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.

8. 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.



9. 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.

10. 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.

NOTE: 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.

CLAN – measuring pause lengths



You may wish to measure the pauses within your data. Pauses are included in the transcript as numerical values placed between parentheses between speaker turns,

```
54 *ROS: and you have written (0.6) five six books↵
55 (0.5)
56 *FRA: yeah
```

or within a speaker's turn.

```
54 *ROS: and you have written (0.6) five six books↵
```

To measure a pause length:

1. locate the silence in the waveform at the bottom of the window
2. mark out the section that corresponds to the silence

```
54 *ROS: and you have written (0.6) five six books↵
55 (0.5)
56 *FRA: yeah↵
```

```
20sep13[E|CHAT] * 57 : W 1:19.837-1:24.301; D 00:00:00.598; C at 1:21.004
```



3. Click on the black bar immediately above the waveform, until you see the same information displayed as you see in the screenshot.

a. Q. What could the various sets of figures (W; D; C) represent?

```
20sep13[E|CHAT] * 57 : W 1:19.837-1:24.301; D 00:00:00.598; C at 1:21.004
```

b. Q. Look at the length of the pause as notated in the transcript. Can you see where that information came from?*

4. The values are usually added to a transcript rounded up or down to a ten of a second (0.2; 0.3; 0.4...etc). Micro-pauses rounded up or down to 0.1 second are usually marked as (.)

Exercise:

Included in the workshop materials is a file called Ex4-PauseLength.

Here, we've removed the pauses from the transcript. Can you locate the pause in the media, and using the black bar at the bottom of this page, measure the pause lengths. Add these to the transcript.

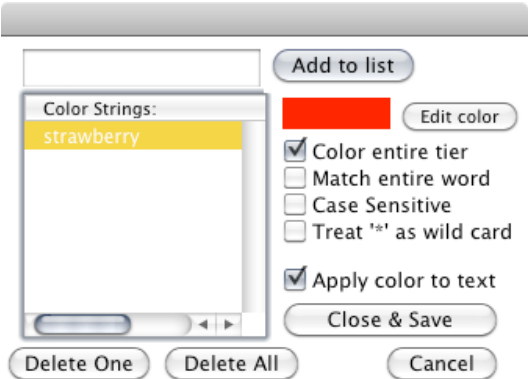
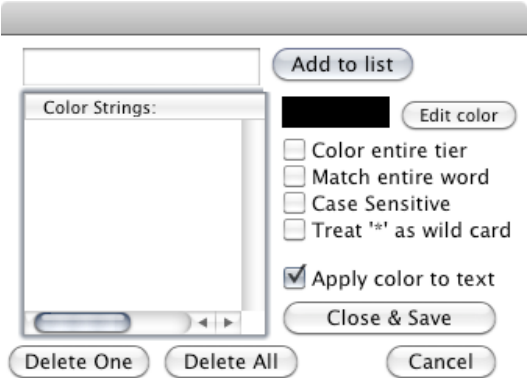
* W indicates the time displayed in the Window; D the Duration of what has been selected; C indicates the placement of the Cursor

CLAN: Colouring keywords



In some cases, it can be useful to use colours to highlight specific parts of a transcript. CLAN allows you to do this through the colouring keywords options. Here’s how:

1. **Open CLAN**
2. Go to **Size/Style** in the top menu and select **Color Keywords**. This opens the box in figure 1.
3. **Type** the string of letters/characters that you want to add some colour to in the blank field at the top.
4. When you’re done, select **Add to list**. Now the string that you’ve entered, in this case ‘strawberry’, appears in the field below (see figure 2).
5. Select the string you have added to 2 and select **Edit color**. Choose a colour you like or otherwise find appropriate, in this case red.
6. **OK**.
7. Use the tick boxes to specify what/how much should be coloured in the colour you’ve selected. In this case the options selected, means that all tiers where the string ‘strawberry’ occurs will appear in red.
8. You can now choose to add more keywords or exit by selecting **Close & Save**.



Advanced option

If you would like to apply the same colour scheme to multiple .cha files, you can do this in a time effective way by following these steps.

1. Add all the colour codes you want to one of the files in your collection.
2. Save the file and close it.
3. Open the folder where the colour coded .cha file is stored, and right click the file to open it with TextEdit (Mac OS) or Notepad (Windows).
4. Copy the @Color words: tier and paste it to all other files .cha files (using TextEdit or Notepad) where you want the same colours to be applied.

```

@Font: CAFont:12:0
@UTF8
@Color words: strawberry 1 65535 0 0
@Begin
@Languages: eng

```



CLAN – Indent

Aligning overlapping talk automatically

You don't need to line up the overlap markers manually. If you put all the second overlaps at the start of the sentence, CLAN will line them up for you automatically.

Note: if you have not added in the overlap markers correctly, the programme will also point these mistakes out when you run this command.

This is how:

1. Open the Command window (**Windows>Command**)
2. Type '**indent**' followed by a space
3. Click on **FILE IN**
4. **Locate** the file you wish the overlaps to align in.
5. Add this file to the right panel.
6. Click **Done**
7. Click **Run**
8. A new file will have been created in the folder, with the extension **indnt.cex**
9. Open this file and you will see that the overlaps have been aligned.
10. Note: at times this doesn't work completely. It's worth checking the right overlaps have been brought into alignment, esp. with multiple overlaps in multiparty interaction.

BEFORE

```
68 *ROS: three ?co-? erph- [i've got it i've got quite a long list here]
69 *FRA: [huh huh huh huh huh huh huh huh]
70 *ROS: thuh huh huh huh
71 *ROS: heh heh .hhh
72 (0.6)
73 *ROS: erm (0.3) [how]
74 *FRA: [i do love] sailing boats
75 *ROS: you do- you love sailing boaf ts]
76 *FRA: [no ?b-] no? just kidding
77 *ROS: [huh huh huh huh huh]
78 *FRA: [huh huh huh huh] huh huh
```

AFTER

```
68 *ROS: three ?co-? erph- [i've got it i've got quite a long list here]
69 *FRA: [huh huh huh huh huh huh huh huh]
70 *ROS: thuh huh huh huh
71 *ROS: heh heh .hhh
72 (0.6)
73 *ROS: erm (0.3) [how]
74 *FRA: [i do love] sailing boats
75 *ROS: you do- you love sailing boaf ts]
76 *FRA: [no ?b-] no? just kidding
77 *ROS: [huh huh huh huh huh]
78 *FRA: [huh huh huh huh] huh huh
```


CLAN – making collections: GEMS



You may want to organize certain sequences in your data into a collection of instances of a particular phenomenon or feature of interest, and to subsequently export all of these instances into a single file.

To make a collection:

1. In the CLAN file, locate the sequence that you want to add to a collection you can search for at a later stage.
2. Add a line before the sequence. Place the Begin GEM header @BG: followed by a tab, and then add the collection name, for example 'other_repair', 'extreme_case'.
3. Create a line after the last line of the sequence you want to mark. Add the End GEM header @EG:, followed by a tab, and again add the name of the collection.

```
23 *ROS: no i
24 i-i was hoping (0.2) i've i've worn them for you basically
25 i was hoping that you 'would'
26 → @BG: laughter_sequence
27 *FRA: [but some] questions tell more about yourself
28 [ʔas they:ʔ]
29 *ROS: Lah::]
30 *FRA: ehuh huh huh huh huh [huh yeah]
31 *ROS: [go on as-] ask me anything then
32 huh [huh heh heh hhh hah hah]
33 *FRA: [h hhhh er:: huhuh what about] the colour red
34 → @EG: laughter_sequence
35 *ROS: .hhhh ah:: now that's an interesting one
```

To search for all the items in a collection:

- a. Open the Commands window (**Windows>Commands**)
- b. Type 'gem **+snameofcollection**' (for example: **+slaughter_sequence**)
- c. Click on **FILE IN** and locate the file(s) you want to search in
- d. Add the file(s) to the right box and click **Done**
- e. Click **Run**
- f. The collection will appear in a new output file (see below). This can be saved as a .cha file for playback.

```
32 *** File "/Users/spencer/Workshop CLAN/Nottingham/02. Linking Software
Workshop Exercises/Ex10-GemCollection.cha": line 64.
33 @BG: laughter_sequence
34 *FRA: i do love sailing boats
35 *ROS: you d- you love sailing boaf'ts]
36 *FRA: [no #b-] no# just kidding
37 *ROS: [huh huh huh huh huh]
38 *FRA: [huh huh huh huh] huh huh
39 (1.1)
40 @EG: laughter_sequence

42 *** File "/Users/spencer/Workshop CLAN/Nottingham/02. Linking Software
Workshop Exercises/JMSH_BLT_BV_20130513_YT_IN.cha": line 10.
43 @BG: laughter_sequence
44 *ROS: i- i think people think that germans eat a lot of mus[sels]
45 *BIR: [no]
46 not at all
47 *ROS: no#
48 huh huh huh huh huh huh
49 *BIR: in cologne but [not]
50 *ROS: [in] cologne
51 *BIR: [but not in germany]
53 @EG: laughter_sequence
```

- g. If the new file contains sequences from different media files, make sure all media files are in the same folder as the output file, should you wish to play each sequence back using the same file.



CLAN – COUNTING GEMs

If you wish to know how many of each GEM you have in your files, you can use the freq command.

This is what you do:

1. Open the Command window
2. Type freq
3. File in> select the file or files you wish to analyze
4. Tier > Include @BG:
5. Done
6. Tier > exclude speaker tiers
7. Done
8. The command should now look like this

```
freq @ +t@BG: -t*
```

9. Run the command
10. You will see that the output counts the tokens included in the @BG tiers:

...

```
4 act-patient
1 and
2 convergence
2 exam-modelling
1 medical-term
2 prefacing
3 repetition
1 safety-netting
6 transition
```

```
-----
29 Total number of different item types used
65 Total number of items (tokens)
0.446 Type/Token ratio
```

Note: If you have not connected separate words in a phrase, it will count them as separate tokens. For example, a GEM named 'extreme case' will output as

```
extreme
case
```

To avoid this, use some linking symbol ('-' or '_') when naming the GEM.

CLAN – hiding and displaying tiers



You may wish to hide some of the tiers in your transcript.

This is very easy to do, and it is equally easy to make them visible again.

To hide a tier:

4. In the CLAN menu bar, select Mode>Hide tiers
(alternatively use the shortcut Esc-4)
5. At the bottom of the screen you will see this:

```
51      so in that case seven books •
52  @Eg:  other_repair
53  *ROS:  erm .hhh a:nd (.)
54  @Bg:  laughter_sequence
55  @Bg:  other_repair
56  *ROS:  but five of them have been translated into english •

20sep13[E|CHAT] 49
: Show, hide or reset tiers? (s/h/r): |
```

6. type 'h' (hide) and hit enter
7. add a tier you wish to hide (for example *ROS or %pic: or @Bg:)
8. hit return
9. the tier should become hidden

To display a tier that has been hidden

1. Follow the same procedure as above but select 's' (show) or 'r' (reset) when prompted.
2. Reset will show all hidden tiers, show will let you select which tiers you would like to display



CLAN – Piping

PIPING in CLAN refers to the facility for directing the output from one analytic tool (for example COMBO or GEM) to another tool for a second analysis. The output from the combined analyses is then collated into a single output file. For example, should you have annotated a batch of transcripts with a particular GEM collection, and would like to conduct further searches within this collection and this collection only, the GEM command can be exported to a second tool, which will conduct an analysis on the particular GEM sequences, before generating an output file.

We will look at this, while referring to the Ex11_PIPING .cha file in the workshop folder.

This is how:

1. In the workshop folder there is a PIPING exercise file. Open it.
2. You see 2 GEM collections: 'laughter_sequence' and 'other_repair'¹. Two of the 'repair' GEMs are embedded within 'laughter' sequences, and one 'repair' GEM is not.
3. We are only interested in the 'repair' GEMs that feature in the 'laughter_sequence' GEM sequences, and are going to generate a file that includes these, and excludes the other one.
4. Open commands, and type
5. gem +slaughter_sequence
6. +d (this directs the programme to export the collection as a CHAT file)
7. File in> choose the file you are subjecting to analysis (Ex13_PIPING)
8. | (the piping symbol)
9. gem +sother_repair²
10. in full, it should look like this

gem +slaughter_sequence +d @ | gem +sother_repair

This basically denotes that an initial analysis will be conducted and will output all the 'laughter_sequence' GEMs, it will furthermore output these as a .cha file (+d), and this will then be piped (|) to a second analysis, where only the GEM sequences marked 'other_repair' will be output.

The eventual file will then only include the 2 'other_repair' sequences that were embedded within the 'laughter_sequence' section, and will exclude the 'other_repair' GEM that was not associated with a 'laughter_sequence' sequence.

See next page for output file. You see that it includes the 2 'other_repair' sequences embedded in the initial 'laughter_sequence' sequence, and has not included the third 'other_repair' GEM. Remember to save this output file as a .cha file if you want to play the data from this file.

EXERCISE

Conduct an analysis on the laughter_sequence GEM collection, specifically looking at how much each participant speaks (hint: the duration of each person's time at talk within these sequences).

¹ It should be noted that the workshop files are for training purposes only, and do not include annotations based on proper analysis

² Note that the programme only searches for the initial standalone word, unless directed not to (see info sheet on COMBO for directions on how to combine words in a search)

```

> gem +slaughter_sequence +d @ | gem +sother_repair
gem +slaughter_sequence +d @
Fri Jan 9 15:29:08 2015
gem (20-Sep-2013) is conducting analyses on:
  ALL speaker tiers
  and ONLY header tiers matching: @BG;; @EG;;
*****
From file </Users/spencer/Workshop CLAN/Nottingham/02. Linking Software Workshop
Exercises/Ex11-Piping.cha>
From file </Users/spencer/Workshop CLAN/Nottingham/02. Linking Software Workshop
Exercises/JMSH_BLT_BV_20130513_YT_IN.cha>
gem +sother_repair
Fri Jan 9 15:29:08 2015
gem (20-Sep-2013) is conducting analyses on:
  ALL speaker tiers
  and ONLY header tiers matching: @BG;; @EG;;
*****
From pipe input
*** File "pipeout": line 12.
@BG: other_repair
*DIM: but five of them have been translated into english
      (0.5)
*ROS: or six
*FRA: well three
*ROS: three co- erph- 'i've got it i've got quite a long list here'
@EG: other_repair
@EG: laughter_sequence

*** File "pipeout": line 26.
@BG: other_repair
*FRA: 'i do love' sailing boats
*ROS: you d- you love sailing boaf'ts'
*FRA: 'no b- no' just kidding
@EG: other_repair
@EG: laughter_sequence

*** File "pipeout": line 36.
@BG: other_repair
*ROS: i- i think people think that germans eat a lot of mus'sels'
*BIR: 'no'
      'no'
*ROS: no'
@EG: other_repair
@EG: laughter_sequence
>

```


CLAN – embedding still images



You may want to add still images to your transcript, for example screenshots of documents the participants are referring to in the talk, or a screenshot of the intonation contours. These will then appear when you reach the selected point in the transcript.

This is how:

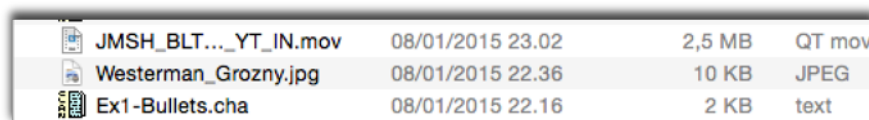
1. Make sure you are not in Sonic Mode (See Mode menu)
2. Insert a blank line into the transcript.

```
29 *ROS: +≈ you travelled an awful lot •
30 which means and you experienced an awful lot≈ •
31 *FRA: |yeah|
32 |
33 *ROS: +≈ you were in yugoslavia and screbenica and so on too •
```

3. In the blank line, add the header %pic: followed by a tab, and (optional) a name for the image

```
30 which means and you experienced an awful lot≈ •
31 *FRA: |yeah|
32 %pic: Westerman in Grozny
33 *ROS: +≈ you were in yugoslavia and screbenica and so on too •
```

4. Add a bullet (**ctrl+i**)
5. Locate the picture (not sure all formats are supported: definitely jpegs though)



6. Click Open File

```
30 which means and you experienced an awful lot≈ •
31 *FRA: |yeah|
32 %pic: Westerman in Grozny •
33 *ROS: +≈ you were in yugoslavia and screbenica and so on too •
```

7. The image is now added to your CLAN file. **Ctrl+click** on the bullet to see the image.
8. When playing the file with **F6**, the image will appear in a separate box when you reach the place where you inserted the file.

Thumbnails– working with gesture annotation



CLAN thumbnail files for multimodal transcription

Transcripts that include annotations of social actions that are not encoded in the verbal modality can be both difficult to produce, and incredibly challenging to read. Eye-gaze, gestures, posture-shifts, features oriented to in the local environment etc may all be relevant to the unfolding interaction, but describing such actions within the main body of the transcript can lead to a loss of ‘readability’.

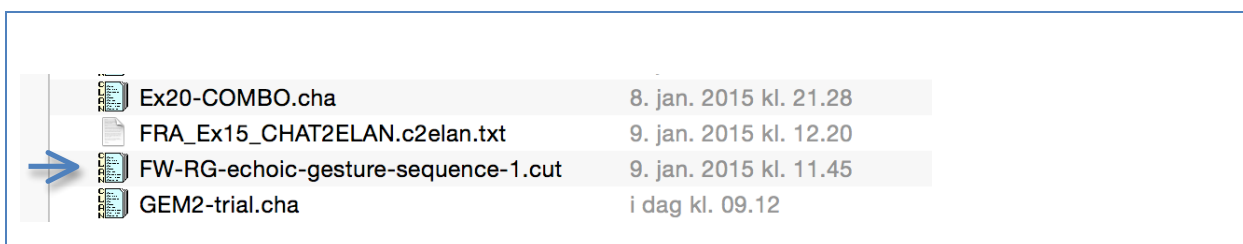
The movie thumbnail feature in CLAN allows you to embed secondary transcription files within the main transcript, which can be activated when a reader/analyst chooses to look at a sequence in greater detail, but without the amount of detail always being visible in the transcript.

This is how you do it.

1. Select your sequence. In your CLAN transcript, find the line/sequence you wish to add a secondary annotation to (for example, a gesture you would like to highlight).

```
30 *ROS: i-i was hoping (0.2) i've i've worn them for you basically
31 i was hoping that you [would]
32 *FRA: [but some] questions tell more about yourself
33 [ʔas they:ʔ]
34 *ROS: Lah::]
35 *FRA: ehuh huh huh huh huh [huh yeah]
36 *ROS: [go on as-] ask me anything then
37 huh [huh heh heh hhh hah hah]
38 *FRA: [h hhhh er:: huhuh what about] the colour red
39 *ROS: .hhhh ah:: now that's an interesting one
40 erm it's definitely my favorite colour ehuh huh but (0.4)
```

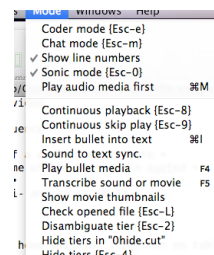
2. Create a .cut file. Start by opening a new file in CLAN,
 - a. give it a name that refers to the sequence you've selected, save it to the same folder
 - b. but save as '.cut' rather than '.cha' (Example 'Griffin iconic.cut')



- c. Add the header @Media, followed by a tab. Then add the name of the video file (without the .mov extension), followed by a comma, a space and 'video' (this should be the same media file as in the main transcript file).

```
1 @Media: JMSH_BLT_FW_20130513_YT_IN, video
2
```


- d. Mode>Transcribe audio or video (or F5). This will play the movie file. Click the main window to stop
- e. Mode>Sonic Mode (or esc+0) will display the waveform
- f. Note: these steps are exactly the same as how you usually open a CLAN file.
- g. On another line, add the header %pic: followed by a Tab and 'entire sequence'



- h. In the waveform, select the sequence you're going to work on, and add it as a bullet to 'entire sequence'

```
1 @Media: JMSH_BLT_FW_20130513_YT_IN, video
2
3 %pic: Entire sequence •
```

- i. Copy and paste the sequence you want to work on, from the main transcript into the .cut file

```
1 @Media: JMSH_BLT_FW_20130513_YT_IN, video
2
3 %pic: Entire sequence •
4 *ROS: i-i was hoping (0.2) i've i've worn them for you basically •
5 i was hoping that you [would] •
6 *FRA: [but some] questions tell more about yourself•
7 [was they:] •
8 *ROS Lah::]
9 *FRA: ehuh huh huh huh huh [huh yeah] •
10 *ROS: [go on as-] ask me anything then •
11 huh [huh heh heh hhh hah hah] •
12 *FRA: [h h h h er:: huhuh what about] the colour red •
13 *ROS: hhhh ah:: now that's an interesting one •
14 erm it's definitely my favorite colour ehuh huh but (0.4) •
```

- j. Note you can also open the bullets with esc+A. This allows you to adjust the bullets manually, for example if you want to ensure the 'Entire sequence' matches the sequence you have copied. Close them again by repeating the command esc+A.

```
3 %pic: Entire sequence •39279_56826•
4 *ROS: i-i was hoping (0.2) i've i've worn them for you basically •39279_42344•
5 i was hoping that you [would] •42355_43632•
6 *FRA: [but some] questions tell more about yourself•
7 [was they:] •43614_46725•
8 *ROS Lah::]
9 *FRA: ehuh huh huh huh huh [huh yeah] •46725_48037•
10 *ROS: [go on as-] ask me anything then •48026_49732•
11 huh [huh heh heh hhh hah hah] •49709_52437•
12 *FRA: [h h h h er:: huhuh what about] the colour red •49674_52809•
13 *ROS: hhhh ah:: now that's an interesting one •52751_54527•
14 erm it's definitely my favorite colour ehuh huh but (0.4) •54511_56826•
```

- k. You are now ready to add further annotations.

- l. Insert a line below the line where you wish to add further annotation, and name the new line according to the type of action you'd like to annotate (%ges: - gesture; %gaz: gaze ; %pos: posture etc.)
- m. Add an identifier at the position you choose at starting point (for example F1 for Frank1)
- n. Select in the waveform the segment that matches the timing of this gesture, and add a bullet to the identifier (Ctrl-I). If you click on the bullet, it should now play the sequence you've selected. Again, you can adjust the bullet manually (esc+A).

```

6      *FRA: [but some] questions tell more about yourself
7      [ʔas they:ʔ]
8      *ROS: lah::]
9      *FRA: ehuh huh huh huh huh [huh yeah]
10     %ges: F1 •

```

- o. Mark the unfolding gesture depending on the conventions you wish to use (for example: preparation; stroke -----, hold +++++, retraction,,,,,,,,,,,,,))

```

6      *FRA: [but some] questions tell more about yourself
7      [ʔas they:ʔ]
8      *ROS: lah::]
9      *FRA: ehuh huh huh huh huh [huh yeah]
10     %ges: F1 • ++++++,,,,,,,,,,

```

- p. Beneath the copied sequence, you can now add a more detailed description of the actions involved in this sequence.
 - i. Start with %pic:, followed by a tab, the identifier (for example F1), and the bullet (these can simply be copy and pasted from above).
 - ii. Then add whichever lines of description you care to. In this example included, they consist of Action, Gaze, Classification, Meaning.

```

29     %pic:      F1 •
30     action:   two handed reach forward, retracted to rest on lap
31     gaz:      to ROS
32     Classification:  deictic
33     meaning:   you

```

- iii. Note: this example has only one gesture described. A sequence can contain as many as you choose. For the others, simply repeat the above procedures.

- q. You have now created your .cut file. Save and return to the main transcription.

3. Embedding the .cut file in the main transcript.

- a. In the main transcript, insert a line immediately prior to the sequence, and give it the header @T: followed by a Tab, and add an identifier (for example, 'deictic', 'ROS gesture', or 'The Guardian').
- b. Add a bullet (Ctrl-I) (Note: make sure you are not in Sonic Mode). A window will pop up to ask you for the file you wish to link to. Choose the .cut file

```

29 @T: FW-RG-echoic-gesture-sequence-1 •
30 *ROS: i-i was hoping (0.2) i've i've worn them for you basically •
31 i was hoping that you 'would' •
32 *FRA: [but some] questions tell more about yourself •
33 [as they:ʔ] •
34 *ROS: Lah::]
35 *FRA: ehuh huh huh huh huh [huh yeah] •
36 *ROS: lgo on as-] ask me anything then •

```

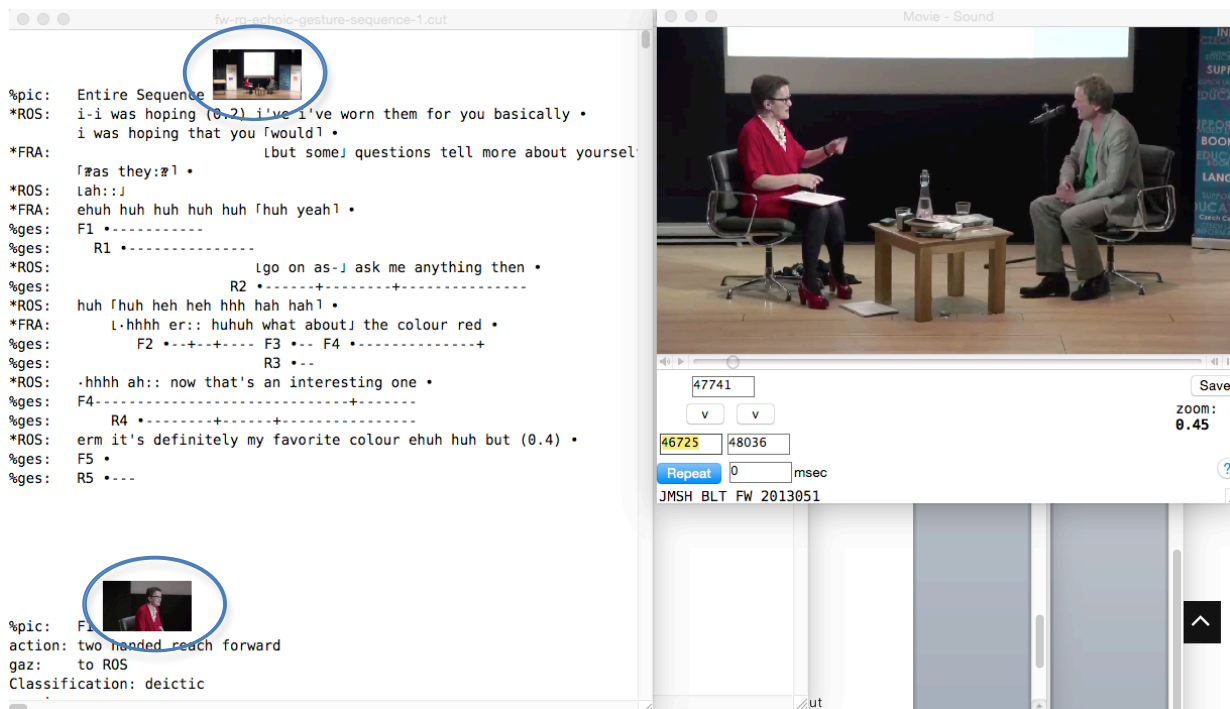
- c. If you would like to check that the bullet now contains a link to the right file, use `esc+A` to open the bullets. Close them with the same command.

```

29 @T: FW-RG-echoic-gesture-sequence-1 •%txt:"fw-rg-echoic-gesture-sequence-1"•
30 *ROS: i-i was hoping (0.2) i've i've worn them for you basically •39279_42344•
31 i was hoping that you 'would' •42355_43632•
32 *FRA: [but some] questions tell more about yourself
33 [as they:ʔ] •43614_46725•
34 *ROS: Lah::]
35 *FRA: ehuh huh huh huh huh [huh yeah] •46725_48037•
36 *ROS: lgo on as-] ask me anything then •48026_49732•

```

- d. You can now Ctrl-Click on the bullet and the .cut file will open in a new window.
- e. Click on the movie thumbnails to see the selected gesture sequences you have annotated. When finished, simply close the .cut file



- f. Note of caution: be careful not to have the @T: thumbnail tiers open when you use the F6 command. This will trigger the .cut files, and this has the habit of crashing the programme. You can still use the F5 command to play the file, without triggering the .cut file. Should you wish to use the F6 command, hide the @T: tiers by using the `Mode>Hide tiers` function (or `Esc+4`). Here you can select which tiers to display or hide.

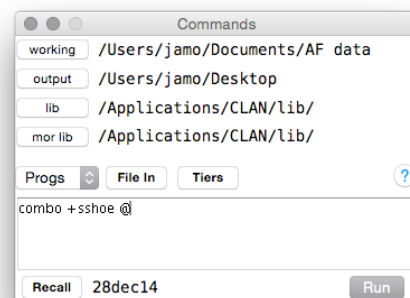
CLAN – searching and coding with COMBO



COMBO is a command that runs from the Commands window in CLAN. It is a very flexible tool that can be used for a variety of purposes. Below we will go through five ways of using COMBO.

1) Searching for strings using COMBO

- 1 Open the Commands window
- 2 Enter COMBO (the name of the command you want to run)
- 3 Press the spacebar and add +s.
- 4 Enter the search string you're interested in immediately after +s, e.g. +sshoe if you want to search for the word 'shoe'
- 5 Use 'File in' to specify the .cha file (or files) on which you want to perform the command
- 6 Save the output file in the same folder as the media file. You'll now be able to play the bullets.



Searching for complex strings

A search string can be a single word, but it can also be a complex string composed of several words and special symbols. The most typical symbols and their meanings are:

Meaning	Symbol
Immediately FOLLOWED by	^
Inclusive OR	+
Logical NOT	!
'Wildcard'	*

- 7 Say you want to search for the word 'fanatic' used with the premodifier 'shoe'. Type the following in the Commands window (and remember to add 'file in'):

combo +sshoe^fanatic @

This will return hits of 'shoe fanatic' (but not just 'fanatic').

- 8 Save the output file in the same folder as the media file. You'll now be able to play the bullets.

2) COMBO search on the basis of pre-established lists

1. Open a new file in CLAN
2. Type the words you want to search for, one word per line.
3. Save the file as a .cut file, e.g. words.cut, in the CLAN library
4. Open the Commands window
5. Build the name of your .cut file into a COMBO search

combo +s@words.cut @

6. Save the output file in the same folder as the media file. You'll now be able to play the bullets.

3) Adding codes with COMBO

Adding codes with COMBO requires two .cut files:

- a. A file with the strings/words you want to add codes to
 - b. A file that contains the coding command
1. The first .cut file is produced as described in section 2) above, points 1.-3.
 2. To produce the second .cut file, open a new file in CLAN, and type a line similar to this:

```
"@words.cut " "$SHOEFAN:" "%cod:"
```

The first part ("@words.cut") specifies the search string. Note: the .cut file can be made part of a complex string, using the symbols in 1).

The second part ("\$SHOEFAN:") specifies the code that will be inserted in the transcript for each hit

The third part ("%cod:") specifies the name of the dependent tier which the code will be inserted on. If the tier is not in the original .cha file, it will be added automatically

3. Save the file as a .cut file, e.g. codingshoe.cut, in the CLAN library
4. Now you're ready to run the actual command. It could look like this:

```
combo +s"@codingshoe.cut" +d4 @
```

+d4 tells CLAN to add codes and tiers to the data file if a string match is found

5. Save the file in the same folder as the media file as QuestionTimeCoded.cha (or something similar). You'll now be able to play the bullets.

4) Cross-tier COMBO

By default, COMBO will only search main tiers. However, you may specify additional tiers to be searched, for instance coding tiers like %cod: introduced above.

1. To search for a string across main tiers and dependent tiers, include the switch +t immediately followed by the name of the independent tier type, e.g. t%cod: Try the following command on the file generated in section 3).

```
combo +s$SHOEFAN* +t%cod: @
```

2. Save the output file in the same folder as the media file. You'll now be able to play the bullets.
3. If you want your hits to be presented with more context, you can specify the size of the 'window' surrounding the hit in the output text by using the switches +w and -w. The following version of the command will give you 4 utterances before (-w4) and 4 utterances after (+w4) the actual hit:

```
combo +s$SHOEFAN* +t%cod: +w4 -w4 @
```

5) Limiting your search with COMBO

Combo allows you to limit your search to particular speakers. That is, you can tell the program/command only to look for strings in specifically designated tiers, for instance all tiers belonging to a particular speaker, e.g. *ROS.

1. To run a limited search you use the +t switch to specify the name of the tiers which the program should focus on in the search, e.g.

```
combo +t*ROS +sshoe @
```

This command returns hits every time the word 'shoe' appears in a tier attributed to the speaker ROS.



CLAN – exporting to ELAN

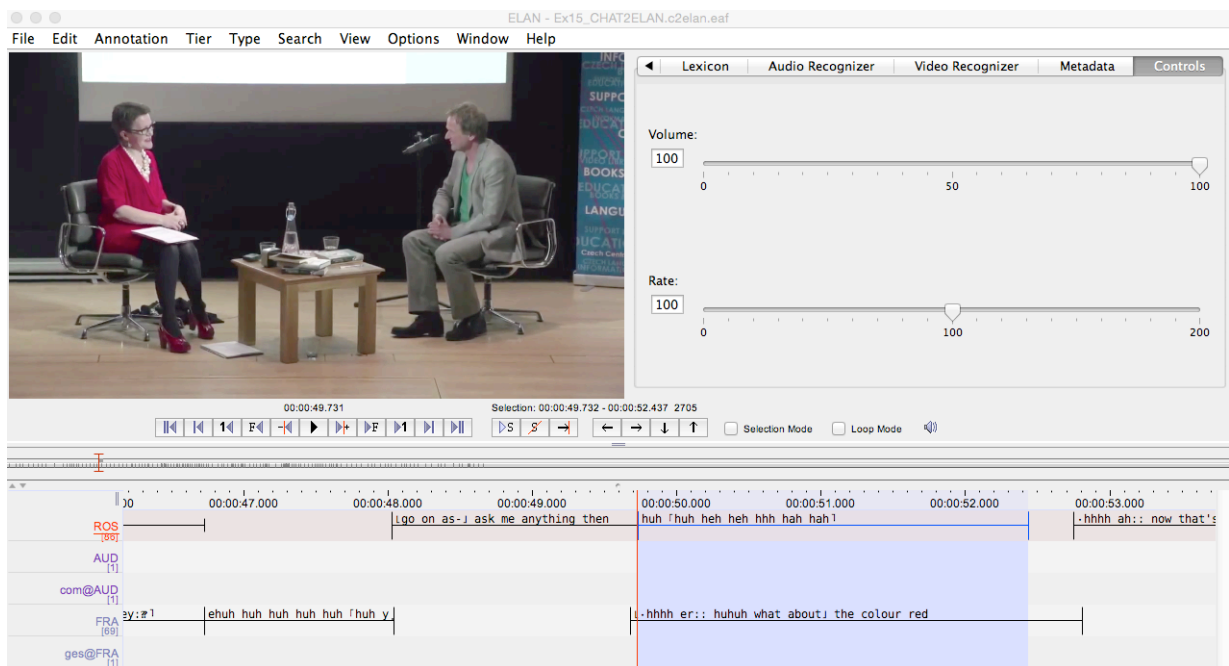
You may want to make the transcript available to be used in ELAN, which allows for other types of data treatment, as well as a different overview (in a partitur type editor).

Here's how you do it:

1. **Open** the Commands window.
2. **Type chat2elan**
3. **File in >** select the required file > **Done**
4. **add +e.mov**

```
chat2elan @ +e.mov
```

5. **Run**
6. In the folder you will now find an ELAN file. **Open it.**
7. The transcript is now available to be used in ELAN



ELAN – adding subtitles to movie files



For presentation purposes, you may wish to include subtitles in the movie file. One way you can do this is by using ELAN and Quicktime Pro.

Here are the steps:

1. In ELAN, select File > Export as > Quicktime text
2. Choose which tier(s) you wish to export as subtitles. Note that you may wish to export different speakers separately, especially where there is overlapping talk in the data.
3. Click on 'Edit font and display settings' and set your preferences for the text colour and size
4. Apply
5. Okay
6. Open the exported file with Quicktime Pro
7. Select all and copy
8. Open the movie file you are adding the subtitles to
9. Edit > Add to movie
10. The subtitles will now be embedded in the movie
11. Should you wish to move them horizontally or vertically:
 - a. Open Window > Show movie properties
 - b. Select the text track
 - c. Add a value into the Offset field (horizontal: left; vertical: right)
 - d. Enter
 - e. The text will have moved to where you have specified.
 - f. Modify until it's where you want it to be.

Should you wish to change the text, you can do this in the original output text file.

1. Open it with a text editing tool (for example TextEdit)
2. You can now modify the text descriptors at the top of the file
3. For relevant codes to do this, see <http://docs.info.apple.com/article.html?artnum=42643-en>

APPENDICES

Appendix 1: List of a selection of special characters for transcription.

You can find these in Windows>Special Characters

	Character	Function	Shortcut (F1+)
1	⌈	top begin overlap	[
2	⌋	top end overlap]
3	⌌	bottom begin overlap	{
4	⌍	bottom end overlap	}
5	≈	≈turn continuation	+
6	≈	≈latching	=
7	°	°softly°	0
8	?	?unsure?	/
9	†	Anonymised names	Alt-0134
10	‰	Name to be anonymised	Alt-0137

Appendix 2: Function key commands for transcription

- F4: play bullet where cursor is placed (or †-leftclick)
- F5: play continuously from bullet where cursor is placed
- †-leftclick on highlighted area of sonic waveform display: play highlighted segment
- triple click on bullet: highlight corresponding section of waveform display. If you have difficulties, leave Sonic Mode (Esc+0) and re-enter (Esc+0)
- F6 (or esc+8): this will play through the transcribed section of the data, highlighting each line as it played in the audio. Place the cursor at the bullet from which one would lie to play the clip.

Appendix 3: Optional Header lines

(Note: these go above the @End line, cf. the order in the header template)

- i. @Situation: Type of interaction (e.g. Interview, Focus group)
- ii. @Date: Date of recording. Use this format: 04-FEB-2010
- iii. @Location: Place of recording
- iv. @Tape Location: The name of the log file, with .log extension.
- v. @Transcriber: Name of transcriber and date of transcription work. Each time a transcriber continues working on the file, he or she should add this information again, separated by a comma.
- vi. @Transcription: Status: 'Partial', 'Full' or 'Checked'
- vii. @Comment: The checker/proof reader must fill in his/her full name, the date of the check and initials (DD-month-YYYY, INI).