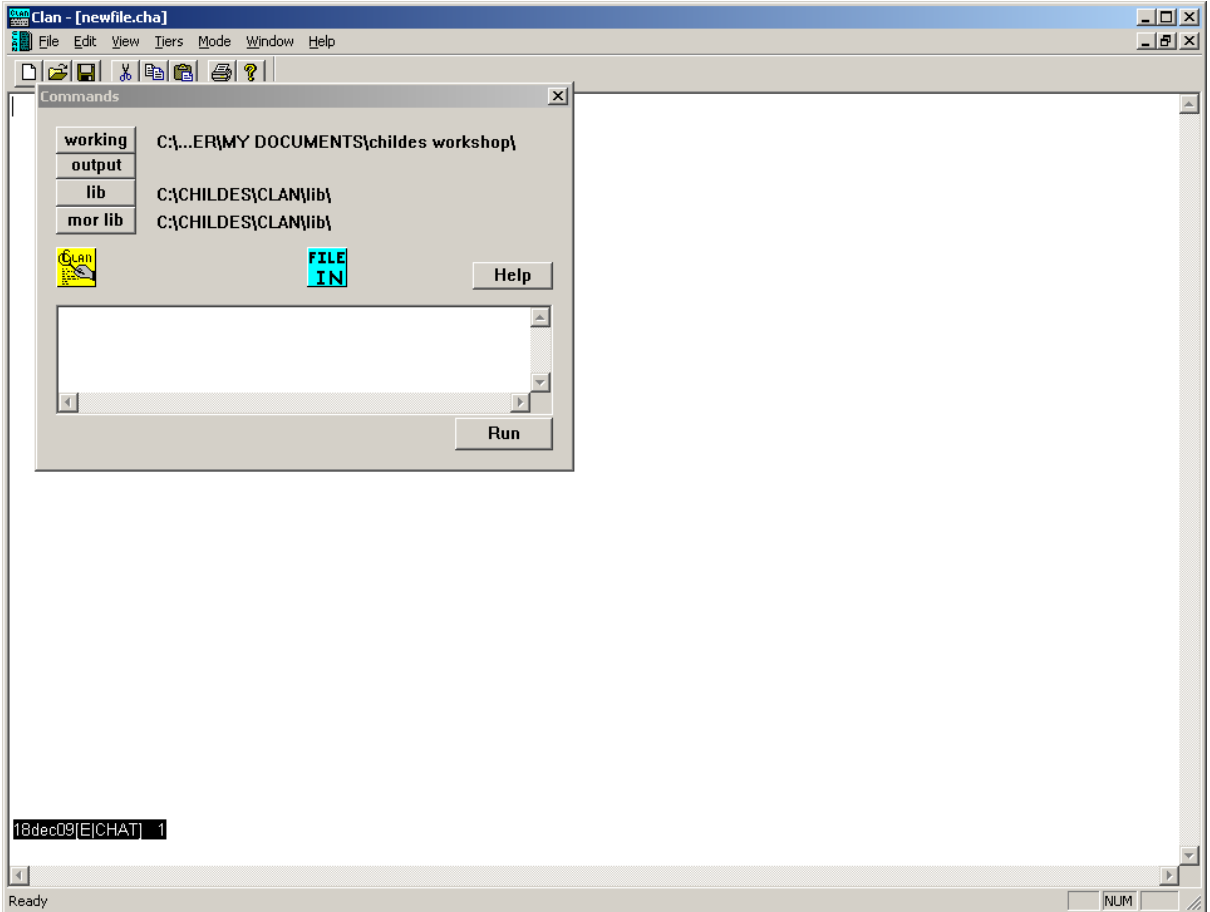
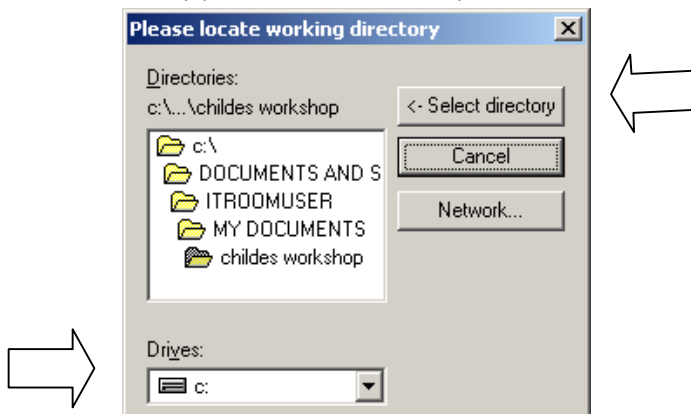


Childes workshop: practical

1. Open the CLAN program. It should look like this:



2. Make sure the working directory is set to wherever you have saved the files you want to analyse. In the Windows screenshot below I have a folder called "childes workshop" in a subdirectory of My Documents.
 - a. To change the working directory press "working".
 - b. Make sure the drive is correct (e.g. c drive) and navigate to the folder where you have saved your files.
 - c. Finally press "select directory"



Info: you need to make sure the working directory is set to where you have saved the files you are going to be working on. In this case we will be working on 5 files from L1 English speaking children. These files are taken from the “Belfast corpus” created by Alison Henry.

Exercises:

1. Open the barb01.cha file.
 - a. Who are the people in the recording?
 - b. What are their names?
 - c. What roles do they have in the recording, e.g. target child?
 - d. What language are they speaking and how can you tell from the headers at the top of the file?
2. Using the command window, calculate the “frequency of the words used” (freq) for the child in barb01.cha.
 - a. What is the most common word she uses? (you will need to look through the list)
 - b. What is the total number of words (tokens) she uses?
 - c. What is the total number of types of words she uses?
 - d. What is the type/token ratio?
3. Calculate the frequencies for all 5 files – hint: replace the filename barb01.cha with *.cha. Fill in the table below with your answers.

	Barb01	Barb04	Barb08	Barb11	Barb15
Most common word					
No. of Types					
No. of tokens					
Type/token ration					

4. How does the type/token ratio change over time? What might this tell us about the child’s development?
5. Calculate the MLU (mean length of utterance) for the child in each file. Again you can calculate this individually per file or you can replace the filename with * (i.e. *.cha)

	Barb01	Barb04	Barb08	Barb11	Barb15
No. of utterances					
No. of morphemes					
Ratio (MLU)					
Standard deviation					

6. What does the change in MLU tell us about this child’s development? Why is the standard deviation important?
7. Find out how many times the child uses a verb ending in –ing in each file.

	Barb01	Barb04	Barb08	Barb11	Barb15
No. of “ing” verbs					

Step by step instructions on how to answer each of the exercises and answers

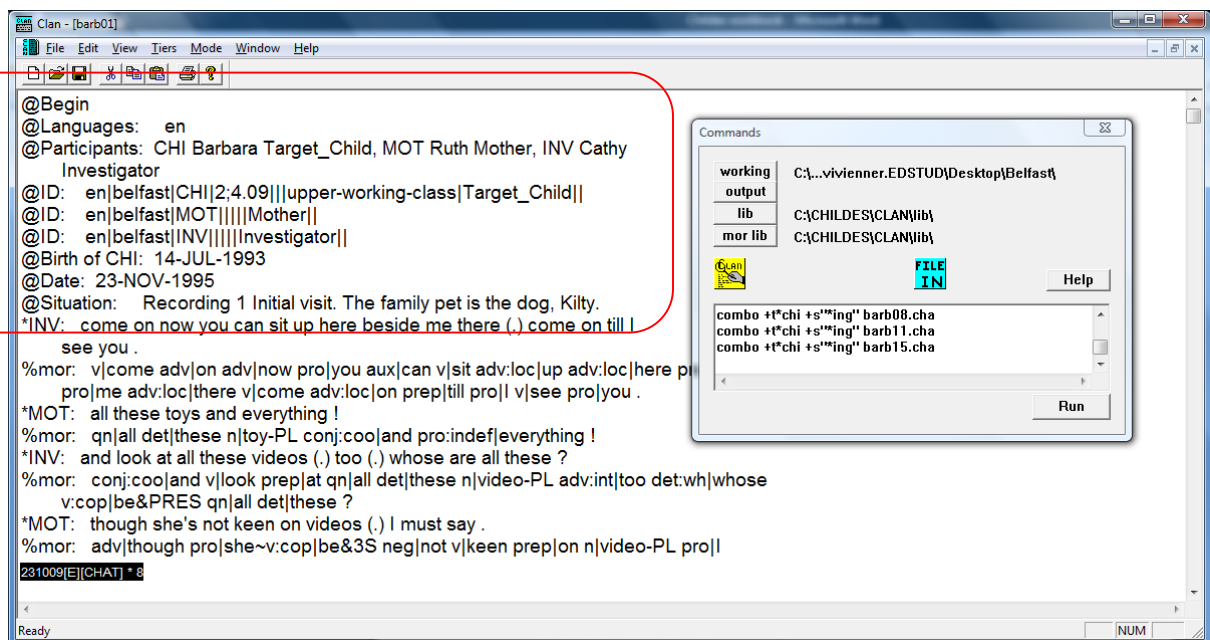
Open the file barb01.cha by going to File>Open and navigating to where you have saved your files.

Question 1

- Who are the people in the recording?
- What are their names?
- What roles do they have in the recording, e.g. target child?
- What language are they speaking and how can you tell from the headers at the top of the file?

Solution

Look at the headers at the beginning of the file.



```
Clan - [barb01]
File Edit View Tiers Mode Window Help
@Begin
@Languages: en
@Participants: CHI Barbara Target_Child, MOT Ruth Mother, INV Cathy
Investigator
@ID: en|belfast|CHI|2;4.09||upper-working-class|Target_Child|
@ID: en|belfast|MOT|||||Mother|
@ID: en|belfast|INV|||||Investigator|
@Birth of CHI: 14-JUL-1993
@Date: 23-NOV-1995
@Situation: Recording 1 Initial visit. The family pet is the dog, Kilty.
*INV: come on now you can sit up here beside me there (.) come on till I
see you .
%mor: v|come adv|on adv|now pro|you aux|can v|sit adv:loc|up adv:loc|here p
pro|me adv:loc|there v|come adv:loc|on prep|till pro|I v|see pro|you .
*MOT: all these toys and everything !
%mor: qn|all det|these n|toy-PL conj:coo|and pro:indef|everything !
*INV: and look at all these videos (.) too (.) whose are all these ?
%mor: conj:coo|and v|look prep|at qn|all det|these n|video-PL adv:int|too det:wh|whose
v:cop|be&PRES qn|all det|these ?
*MOT: though she's not keen on videos (.) I must say .
%mor: adv|though pro|she~v:cop|be&3S neg|not v|keen prep|on n|video-PL pro|
231009[E][CHAT] * 8
```

The screenshot shows the CLAN software interface. The main window displays the headers of a file named 'barb01.cha'. A red circle highlights the header information, which includes the beginning of the file, the language (en), the participants (CHI Barbara Target_Child, MOT Ruth Mother, INV Cathy Investigator), and the birth date of the child (14-JUL-1993). A 'Commands' dialog box is also visible, showing the working output path (C:\...viviennr.EDSTUD\Desktop\Belfast\), the lib path (C:\CHILDES\CLAN\lib\), and the mor lib path (C:\CHILDES\CLAN\mor lib\). The dialog box also shows a list of files: 'combo +t*chi +s*ing* barb08.cha', 'combo +t*chi +s*ing* barb11.cha', and 'combo +t*chi +s*ing* barb15.cha'. The 'Run' button is visible at the bottom of the dialog box.

The third line down begins “@Participants”. This gives you the details of the participants in the recording. In this case there are three:

- Barbara, who is the target child and coded as CHI
- Ruth, the mother and coded as MOT
- Cathy, the investigator (or researcher) who is coded as INV

Following the @Participants line, there are 3 @ID lines. These lines give us further information on each of the participants. For example, it gives us the language, name of corpus, ID code, age, socio-economic class and label, e.g. target_child. Each item of information is separated by a |. Not all the information has been given for each person. It is up to the person making the corpus to decide what information is relevant. In this case, the ages of the mother and the investigator are not relevant.

Question 2: Using the command window, calculate the “frequency of the words used” (freq) for the child in barb01.cha.

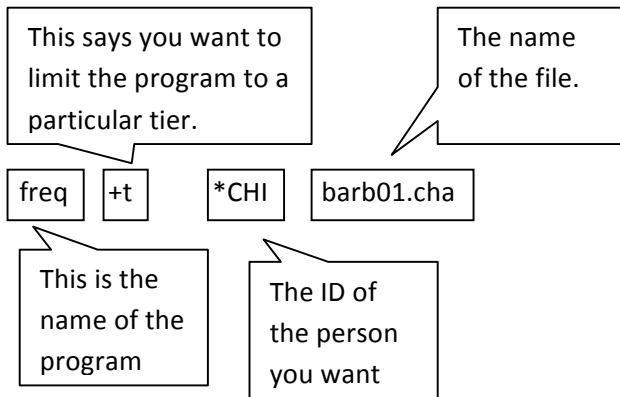
- a. What is the most common word she uses? (you will need to look through the list)
- b. What is the total number of words (tokens) she uses?
- c. What is the total number of types of words she uses?
- d. What is the type/token ratio?

Solution

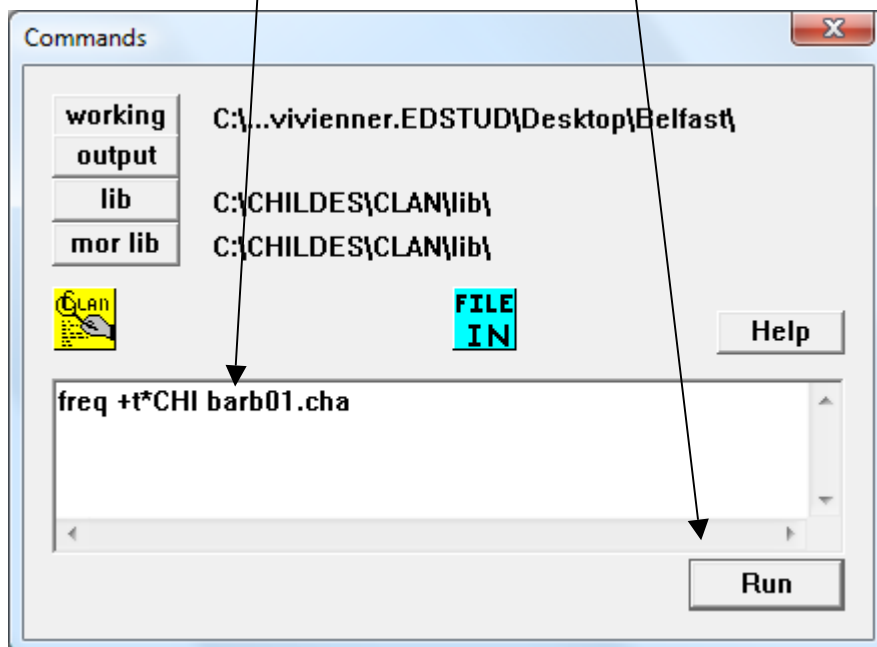
In the command window, you need to type the command

freq +t*CHI barb01.cha

This command can be broken down into the following sections:



Type the command here. You then need to press « Run ».



You will then see a new window appear called “Clan Output”. In this window you will see a list of words and then the following:

149 Total number of different word types used
411 Total number of words (tokens)
0.363 Type/Token ratio

This is the information you need to answer the questions. You should also look through the list of words. Before each word is a number. This number gives the number of times that word is used in the file. Find the one with the highest number – i.e the most commonly used word.

Answers:

- a) What’s (used 40 times)
- b) Tokens = 411
- c) Types = 149
- d) Type/Token ratio (TTR) = 0.363

Question 3: Calculate the frequencies for all 5 files – hint: replace the filename barb01.cha with *.cha. Fill in the table below with your answers.

Solution

Follow the steps as above but this time instead of running the program for each file individually, replace the filename “barb01” with a *.

Command: freq +t*CHI *.cha

	Barb01	Barb04	Barb08	Barb11	Barb15
Most common word	What’s	I	I	the	I
No. of Types	149	279	155	309	107
No. of tokens	411	1158	550	1110	289
Type/token ration	0.363	0.241	0.282	0.278	0.370

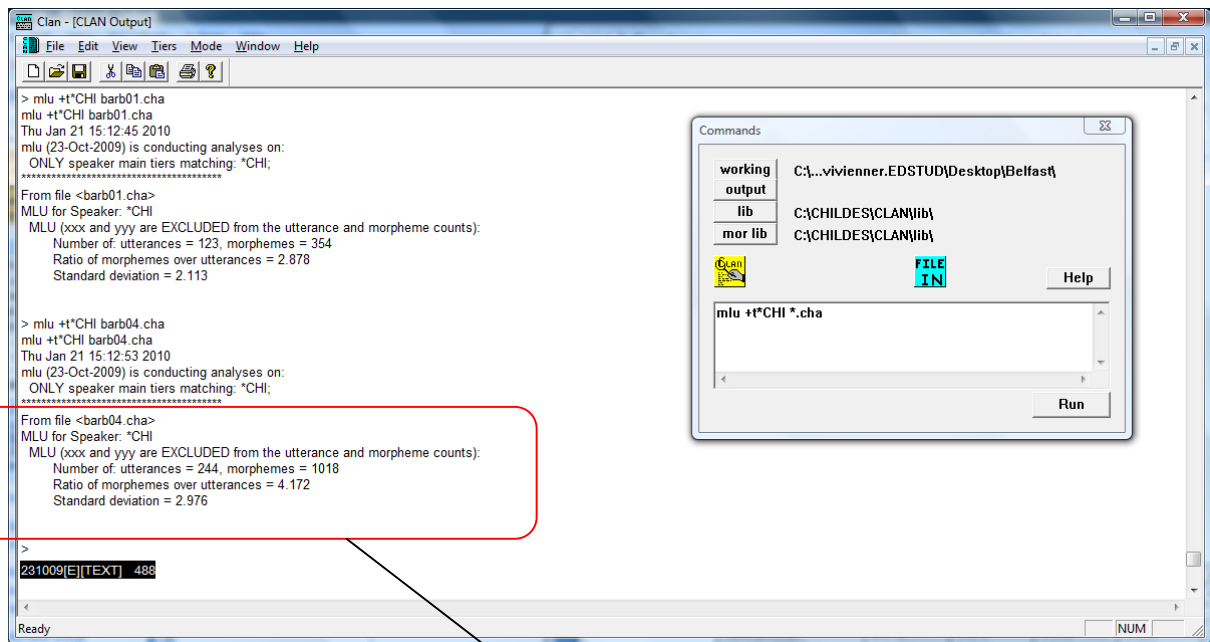
Question 5: Calculate the MLU (mean length of utterance) for the child in each file. Again you can calculate this individually per file or you can replace the filename with * (i.e. *.cha)

Solution

Type in the command mlu +t*CHI *.cha in the command window. Press “run”.

The following “Clan Output” will appear (see next page).

The relevant information for each file is given separately, as is shown for in the highlighted box for barb04.cha



	Barb01	Barb04	Barb08	Barb11	Barb15
No. of utterances	123	244	154	286	65
No. of morphemes	235	1018	514	938	289
Ratio (MLU)	2.878	4.172	3.338	3.280	4.446
Standard deviation	2.113	2.976	2.508	2.527	3.429

Question 7: Find out how many times the child uses a verb ending in -ing in each file. Looking at the example sentences, how many different verbs ending in -ing does the child use in each recording.

Solution

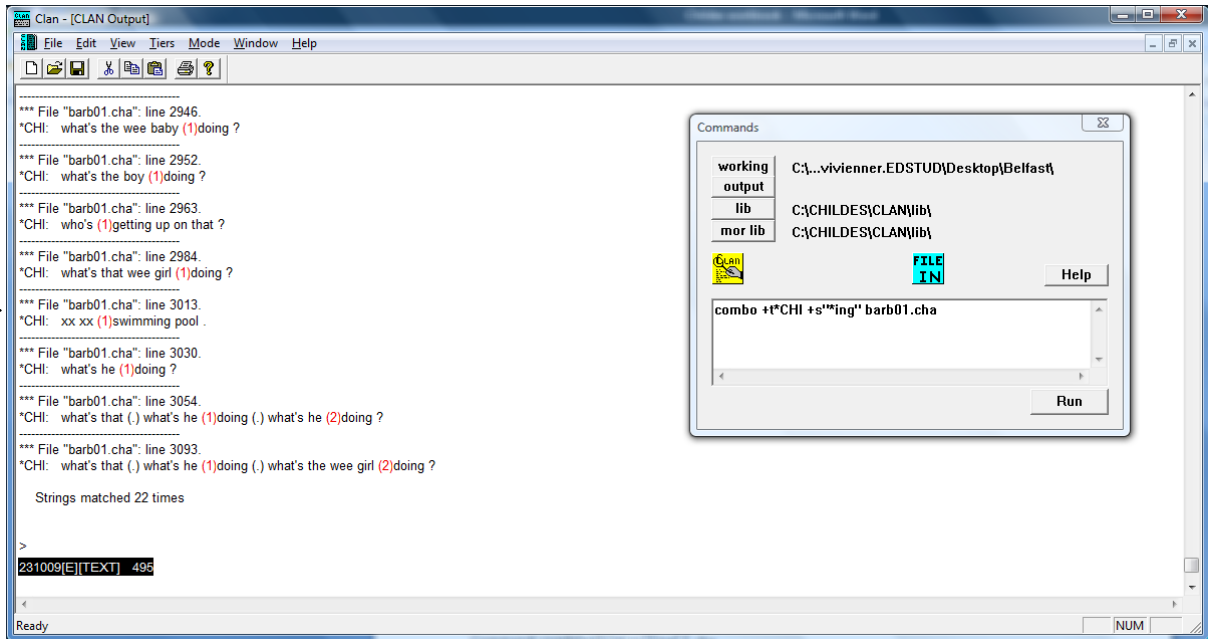
To answer this question there are two ways of doing this. You can either search the main tier (*CHI), or as these files have been morpho-syntactically tagged, you can search the %mor tier.

Option 1: search the main tier (*CHI).

You need to use the “combo” command. You want to find all words ending in “ing” so you search for “*ing” in all the files so you type *.cha

Command: `combo +t*CHI +s"*ing" *.cha`

This command will then open a Clan Output window and give you a list of all the utterances containing “*ing” and then a total number of strings for each file.



However, as you can see. One of the disadvantages of searching this way is that it gives you words like “swimming pool”. We probably don’t want to count these. We can manually go through the 22 items and discount anything non-verbal. For example, in barb01.cha we will want to discount, “a ring”, “nothing” and “swimming pool”.

	Barb01	Barb04	Barb08	Barb11	Barb15
No. of “ing” verbs	19	26	12	18	7

Option 2: search the mor tier (tagged tier)

Many transcripts available on CHILDES and TALKBANK have been morphologically tagged for easier analysis.

You will still use the combo command but you need to tell the computer to search on the %mor tier instead of the main tier and you need to tell it what structure to look for. In this example, you would use the following command:

Command: `combo +s"part|*PROG" +t*CHI +t%mor *.cha`

This command will only look for participles (part) that are also coded for progressive (PROG).