A Clinician's Complete Guide to CLAN and PRAAT

Nan Bernstein Ratner
University of Maryland

Shelley B. Brundage
George Washington University

Last update: 8/28/15

Introduction: WHY you want to learn how to use these programs.

We have all been taught that a speech and language sample is the "gold standard" for evaluating our cases. But you know how tedious language sampling and fluency counts can be, from transcribing to calculating usable clinical scores. We think the last time most clinicians actually DID a language sample was in graduate school. Most fluency specialists, according to surveys we have done, reliably collect a spoken language sample, transcribe, and then code for different fluency types. However, then you can't typically use this sample for much else.

This manual will teach you to use CHAT and CLAN, which will, using a SINGLE TRANSCRIPT:

- Speed up your accurate transcription of data by about ten-fold;
- Link your transcript to the actual audio or video file you made in the clinic;
- For language analysis, automatically (!!!) compute clinical measures of interest, such as MLU, TTR, DSS, Brown's morphemes (for children), and a host of other values (WITHOUT MAKING MATH ERRORS!)
- For fluency analysis, automatically analyze the frequency of different types of disfluency behaviors and compute the percent stuttered words
- Allow you to link to free acoustics software DIRECTLY from the transcript, to calculate values such as speech rate and acoustical features of targeted words or utterances
- Create your OWN targeted analyses of your client's data
- Grammatically tag and analyze data from more than a dozen languages other than English
- Output to PHON for phonological analysis

These programs are open source software available for free, with excellent support from a major US government-supported team that built this wonderful utility, and a list of community users around the world. The three programs we describe in this manual are:

CHAT and CLAN
CHAT and CLAN are a part of CHILDES (CHIld Language Data Exchange System), which provides tools for studying conversational interactions, as well as serving as a repository for language corpora from around the world. CHAT a software program that is used to transcribe sound files using a standard set of rules (commonly called "CHAT format"). CLAN is a data analysis program and is used to analyze transcriptions that were transcribed in CHAT format. (For more information, go to http://childes.psy.cmu.edu/)

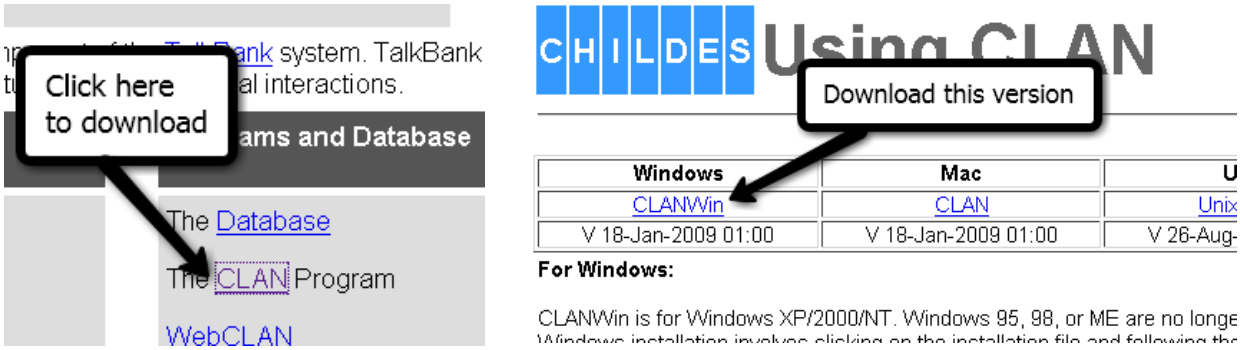
PRAAT
Praat is a freely available program, developed by Dutch researchers, to perform acoustical analysis (see more further in this guide). It intersects directly with CLAN, via a command under MODE menu: send to sound analyzer drop down menu.

Getting Started: Computer setup

First, turn on your computer ☺ and be connected to the Internet.

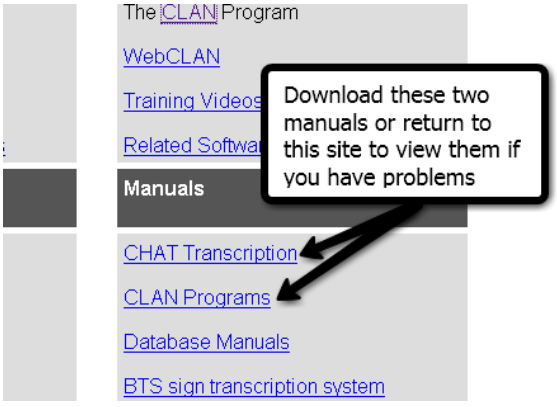
DOWNLOAD CLAN
To download CLAN, go to go to http://childes.psy.cmu.edu/CLAN/.
When you download CLAN you get CHAT along with it. You do not need to download them separately. PC users choose "CLANwin" and Mac users choose 'CLAN." (We doubt any of you are Unix users, but they have that version, too).

This is what it looks like for PC users…



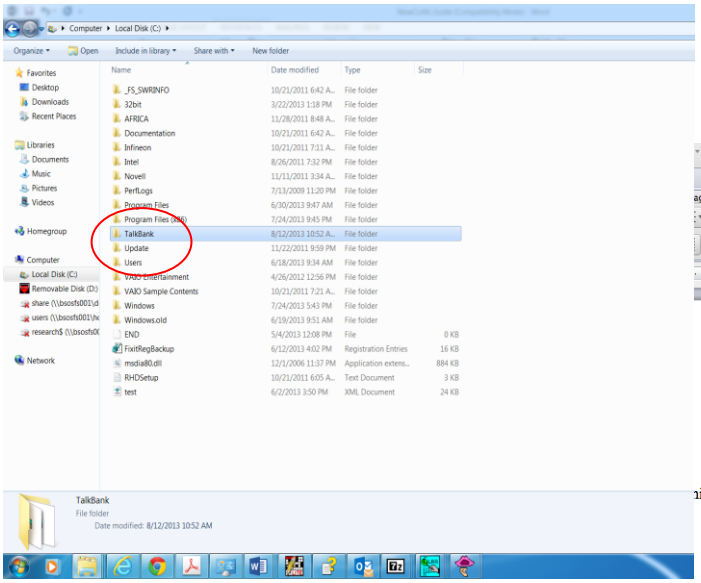
There are official **user manuals** for both CHAT and CLAN. *You should also download the "CHAT transcription system" manual and the "CLAN Programs manual" at http://childes.psy.cmu.edu/.* These manuals are long and technical, so don't print them. Instead, we recommend saving the CHAT manual and the CLAN manual to your desktop. Importantly, these manuals are searchable by keyword, which may be useful later.

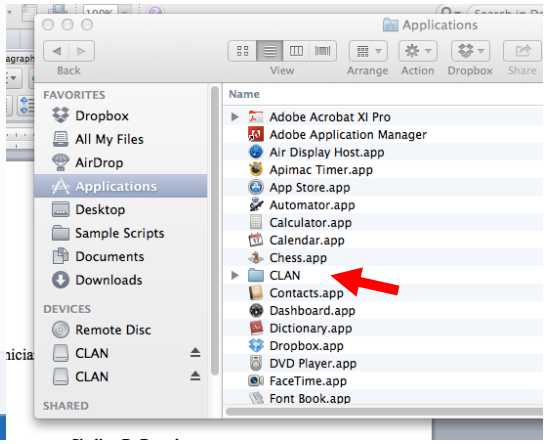
The manual you are reading right now should contain most of what you'll need to know, but if you have any questions, the manuals described above from this website are much more comprehensive!



!!!! One last important download:
Please be sure to download a new version of Quicktime each time you install CLAN: www.apple.om/quicktime/download/. This is crucial to audio and video linking of files.

KNOWING YOUR COMPUTER'S FILE ORGANIZATION
It's kind of surprising, but many folks have gotten so used to working from flash drives or the desktop that they don't really know where files are stored on their computer. The default directory that gets built when CLAN is installed is called TalkBank.[1] It should be located on your C: drive, perhaps like this, on a PC:

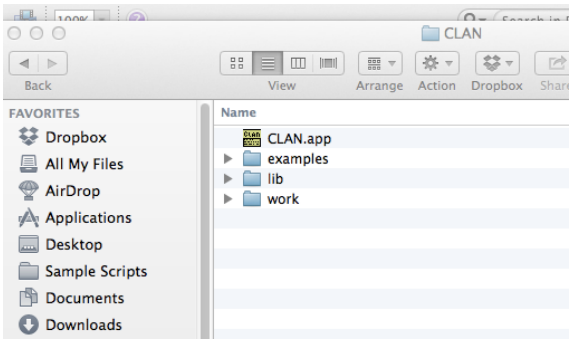




On a MAC, open your FINDER, choose APPLICATIONS, and then you should see CLAN in the list. Double click the CLAN folder and you will see CLAN.app; click to launch.

The Talkbank directory (PCs) and CLAN folder (Macs) have subfolders: the main ones you should get to know are CLAN, LIB, and MOR. To make your life easier, you should also put things you are working on in the subfolder called WORK. Store your audios and videos there, and the program will easily find them when you try to transcribe things.[2] Below is a screenshot of the CLAN folder on a Mac, showing the CLAN application, LIB, WORK, and Examples subfolders

[1] If you are a prior CLAN user, you need to rename your old CHILDES folder as TalkBank, or you will wind up with competing versions of CLAN on your computer. If you can't find a CHILDES folder, never mind ☺
[2] If you need a tutorial on managing files on your computer, go to http://windows.microsoft.com/en-us/windows-vista/working-with-files-and-folders for Windows and http://www.apple.com/findouthow/mac/#gohome for Mac users.



AUDIO AND VIDEO FILES

CHAT (the transcription rule system/program in CLAN) allows you to link a transcript directly to audio or video. This speeds up transcription a LOT, and improves its accuracy. Only a few programs actually use the audio or video for analysis; the primary one is exporting an utterance to PRAAT to do acoustical analysis of what you've recorded. Clinicians may want to use PRAAT to analyze features such as speech rate, articulation of individual words, prosodic contour or range for looking at prosody in clients, as just one example.

For fluency cases, being able to review a transcript and see/hear what the client was actually doing is immensely helpful in being able to accurately count and locate disfluency behaviors. It also serves as a permanent record of what the client was doing and saying that can be maintained and compared over time by multiple listeners, SLPs, researchers, etc.

We can't demonstrate one amazing function of a linked transcript in this manual, but you will be able to see it when you follow all the instructions in this short guide. As you play a transcript in CLAN, it will show you the highlighted transcript lines and you will see/hear that segment of the sample at the same time. Because of the audio-video linkage, you can go directly to any given point in the transcript and listen to it.

The most commonly used audio and video formats that are supported by CLAN are .wav, .mov and .mp3. *Although CLAN supports virtually all audio and video formats, it does not recognize files written to DVD.* If you wind up needing to transcribe a DVD file, go to the web and find one of the many free audio editing programs that convert DVD to other formats, such as Audacity (http://audacity.sourceforge.net/). If you know that you will be doing acoustical analysis of your samples using the PRAAT interface, be sure that your media works with PRAAT (it is a bit fussier).

How to Transcribe In CHAT

Everything we will talk about below is provided in a step by step in a TUTORIAL that uses practice materials, available on the CHILDES web site: http://childes.psy.cmu.edu/ - go to PROGRAMS, then TUTORIAL, and download the zip file. However, the tutorial starts by bulleting already transcribed materials; we find it more useful to bullet the audio or video first, and then, only if you are a perfectionist, re-bullet the file later to make it align beautifully.

** Sometimes, when you open CHAT and attempt to link it to your audio/video files, you get *an error message saying CLAN cannot locate the media file*. If this happens, try the following, in this order:

1. Make sure that the media file name matches the name listed in @Media header in your CHAT transcript
2. Verify that the media file name does NOT have any spaces, or any non-English characters in the filename or in the pathway to it.
3. Make sure the ChAT file and the media associated with it are in the same directory, or that the media (.mov or audio only) is in the Media folder.
4. Make sure that the audio or video file is REALLY an mp3, .mov, etc. You cannot make a file into a different file simply by adding .mov to the end of a .wmv file, any more than you can make a .doc into an .xls simply by renaming it!
5. Restart CLAN
6. Restart your computer
7. Re-download CLAN
8. Make sure Quicktime is installed and working (try it on a non-CLAN file). If not sure, download a new version
9. Note that on PCs, file extension are hidden by default. If the file extension is missing, then you might be tempted to add it. For example, file "Shaggydog.mov" will look on PC desktop like "Shaggydog" only, because extension ".mov" is hidden. If you try to correct this by adding extension ".mov" to "Shaggydog" the visible media file name will be "Shaggydog.mov", but the actual name will be "Shaggydog.mov.mov". CLAN will not be able to find this file.

Now back to the main topic – HOW TO TRANSCRIBE. All CHAT transcripts have certain characteristics: Headers, Main Lines, and Dependent Tiers. Each of these is discussed below.

HEADERS

Headers must be present at the top of every file before you can start transcribing. You actually need to write a tiny bit into a file BEFORE you can start transcribing. To open a new, empty CHAT file, double click the CLAN icon. When the program opens, you will see a command window; this window is used for analysis, not for transcription, so you can close it. Then go to the File menu and click on New. This will give you a blank CHAT file window (default name is newfile.cha) where you can begin transcription. **YOU MUST TRANSCRIBE IN CLAN, NOT IN MS-WORD!!!**

Type in the following headers that are needed to start a file:

@Begin
@Languages: eng (for English, use other codes for other languages; see the full manual)
@Participants: CHI Kiddo Child (first word is Tier name, second word is actual identifier/name of your speaker, and third is "role" or function in the interaction (e.g., child, therapist, father), separate speakers by a comma (see next page for example)
@Media: filename, audio OR video [example: @Media: johnDdx, video] (*leave off the file extension*, CLAN knows which extensions are audio and video.
@End

Save this file, and be sure it saves with .cha as the file extension (last 3 letters after the period. Be sure there are NO spaces, periods, commas, or other punctuation in the part BEFORE .cha. Good examples are ARsession1.cha, Kid5age24mos.cha, JDbaselineflu.cha etc. It will typically store to whatever you marked as your Work directory in CLAN.

Basic Rules for Headers

1. Every line must end with a carriage return (oops, showing our ages here, the Enter or Return key; carriage returns are for typewriters; ask your parents what a typewriter is).

2. The first line in the file must be an @Begin header line.

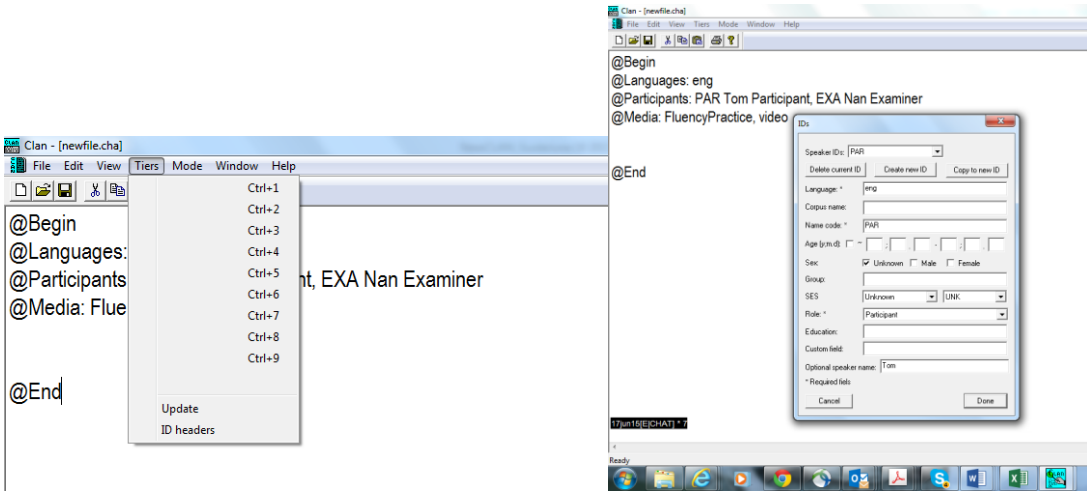
3. The second line in the file must be an @Languages header line, such as "eng" for English.

4. The third line must be an @Participants header line listing three-letter codes for each participant, the participant's name, and the participant's role.

5. The next few lines are a set of @ID headers providing further details for each speaker. YOU DO NOT NEED TO TYPE THESE IN.
These will be inserted automatically for you when you use the pull down option under Tiers (top menu) and select ID Headers, after you have listed all the participants. In this section, you can provide information about the speakers in your file. When you click DONE, it automatically inserts these lines, and adds this information. If you have more than one speaker, you must update each speaker's information.

Note: You can pull this window down at any time, and it can act as a short set of case history information, such as age, gender, diagnosis, and other notes. When you do it, some of it will then reflect in file headers when you open your CHAT file.

Another note: If you keeping track of many clients or subjects, many of the CLAN programs will output results in a spreadsheet. If you have updated information in the ID headers, these will print out as column variables, very handy for tracking client progress, examining patterns on your caseload or for study statistics.



6. IMPORTANT: Note that the @Media header tells the program which audio file you will be transcribing. They must match EXACTLY. If the title of the audio file does not exactly match

the one that you write in the header, then the program will not be able to find the audio. You don't need to put the file extension in the name; CLAN knows that .mp3 or .wav is audio, .mov is video, etc. Your only options for media are audio or video.

7. The last line in the file (after all the main lines and dependent tiers) must be an @End header line.

MAIN LINES

Lines beginning with * indicate what was actually said. These are called "main lines" or "main tiers". There are some basic rules for coding main lines.

1. Each main line should code one and only one utterance. When a speaker produces several utterances in a row, code each with a new main line.

2. After the asterisk on the main line comes a three-letter code in upper case letters for the participant who was the speaker of the utterance being coded. The three-letter code is followed by a colon and then **a tab (not spaces)**.

3. What was actually said is transcribed starting in the ninth space on the line (using TAB in CLAN brings you automatically to this position. REMINDER: Use a tab, not space bar, after each colon.

4. CHAT recognizes a variety of abbreviations for the persons in the audio/video. However, to make maximal use of comparison databases being built that will allow you to compare a child or an adult with speech or language disorder to typical performance, WE SUGGEST THAT YOU ONLY USE *CHI FOR CHILD CLIENTS/STUDY PARTICIPANTS, AND *PAR FOR ANY ADULT CLIENT/PATIENT/STUDY PARTICIPANT. You can choose other three letter codes, but you will lose a lot of the wonderful science you can refer to afterwards if you do that. So, try to listen to this advice! Other typical abbreviations used in the system are MOT for mother and FAT for father and SLP for speech-language pathologist.

5. Utterances should end with an "utterance terminator". The basic utterance terminators are the period, the exclamation mark, and the question mark. The other common ones are used to mark incomplete or trailing off utterances (see below).

6. DO NOT USE commas. The programs don't use them for anything.

7. Try to use upper case letters only for proper nouns and the word "I." If you find that hard, no worry, but CLAN will treat "We" and "we" as different words (unless you pick a CLAN option to ignore case).

8. Unintelligible words with an unclear phonetic shape should be transcribed as xxx.

9. Incomplete words can be written with the omitted material in parentheses, as in (be)cause and (a)bout.

10. Coding fluency: Codes are in APPENDIX 3 of this manual. Fluency is typically coded on the main line [you can add fluency codes to any dependent tier; ask us if you have special needs, such as tagging part of speech that was stuttered].

    We strongly advise both clinicians and researchers to transcribe the sample first for content, and save it. We then suggest you copy it with a new filename, (e.g., copy Tom.cha with a name such as Tomflu.cha, and do your fluency coding on that version). While all of the fluency codes we provide pass CHECK in CHAT, some have been known to complicate the more sophisticated language analyses, or import/export to PHON or SALT. So we suggest that you simply save the file with two different filenames, and use one for fluency analysis and the other for language analyses. They will both still be linked to the same video or audio – you don't need two copies of them.

DEPENDENT TIERS

Dependent tiers are optional additions to the transcript. These tiers begin with the % symbol and can contain codes and commentary regarding what was said in the Main Tier directly above it. There are some basic rules for coding dependent tiers.

1. The % symbol is followed by a three-letter code in lowercase letters for the dependent tier type, such as "pho" for phonology; a colon; and then a tab. The text of the dependent tier begins after the tab.
2. There are a number of pre-suggested dependent tiers for various purposes; consult the CHAT manual for ideas. *You can also insert your own, if you want*. But this will require you to adjust a file to make the transcript pass CHECK, so this is an *advanced topic* (see the end of this manual for advanced topics). **Some programs insert tiers**. For instance, the program MOR that we will describe that runs **automatically** to analyze the grammar in your transcript will insert a %mor dependent tier for you when you run it.

Here are two samples that illustrate headers, main lines, dependent tiers and basic rules described above.

@Begin
@Languages: eng
@Participants: CHI Ross Child, FAT Brian Father
@ID: eng|macwhinney|CHI|2;10.10||||Target_Child|||
@ID: eng|macwhinney|FAT|35;2.||||Target_Child|||
@Media: Ross, audio
*CHI: why isn't Mommy coming?
%com: Mother usually picks Ross up around 4 PM.
*FAT: don't worry.
*FAT: she'll be here soon.
*CHI: good.
@End

(Please note that we didn't start each utterance with a capital letter)

Here is an example from an adult aphasia patient who is in an Aphasia Bank directory of cases called "Adler"(it has some other codes in it, but dis-regard them for now):

@Begin
@Languages: eng
@Participants: PAR adler02a Participant, INV Investigator
@ID: eng|Adler|PAR|69;9.|male|Conduction|adler02a|Participant||74.9|
@ID: eng|Adler|INV||||adler02a|Investigator|||
@Media: adler02a, video
@G: Speech
*INV: I am gonna [: going to] be asking you to do some talking .
*INV: how do you think your speech is these days ?
*PAR: not good . [+ gram]
*PAR: &=laughs &=head:shake not good . [+ gram]
*PAR: &uh <I'm not> [/] I'm not good a(t) this .
@End

SAVING THE .cha FILE AND STARTING THE FULL TRANSCRIPT

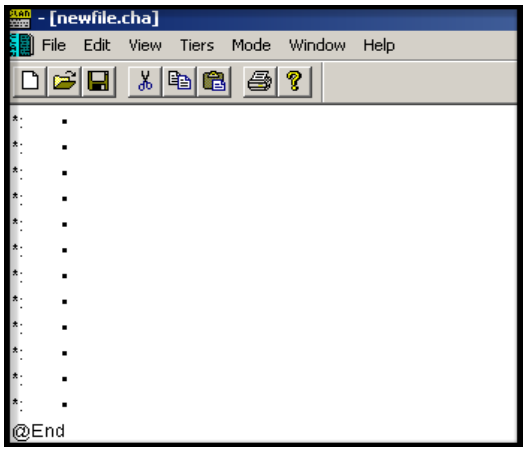
You will not be able to start transcribing using linked audio or video until you have saved the file to your computer with a complete set of header information. Remember to save the file with the exact same name as the audio file except with .cha as the extension. .cha stands for CHAT file format (e.g., NBR.cha, and NBR.mov, or SBB.cha and SBB.wav). The transcript needs to be saved in the same folder as the audio file.

LINKING TEXT TO AUDIO OR VIDEO

This is the truly wonderful aspect of CHAT/CLAN (SALT cannot do this, sorry). Linking speeds up transcription (up to 3 or 4 times, easily, and perhaps much more), increases transcription accuracy and allows an infinite number of reviews or analyses of your session. Now you are ready to do this. Read all the steps below before starting this part of the process.

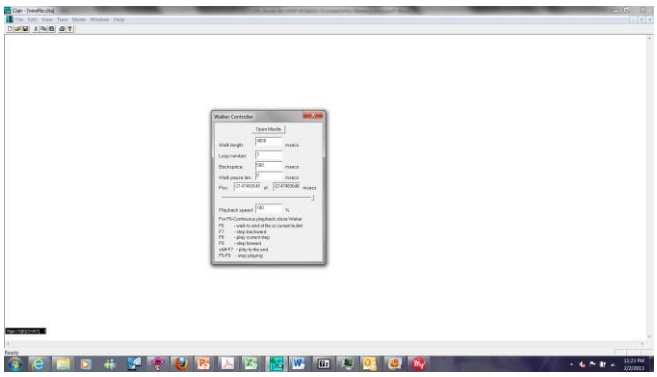
- Open CLAN and open your "starter" file that you just formatted.
- Close the commands window (Click the X in the upper right of box)
- Under the MODE drop down window, select transcribe audio or video OR press the F5 key to begin linking
- In the window that pops up, locate the file to which you wish to link. If you get an error that the program can't locate the file, make sure you have stored it in the same folder (preferably your WORK folder) inside CLAN that you have your starter file in. That is the usual problem.
- When you click on the necessary file, the recording will begin to play
- Each time you hear a silence or natural break, push the space bar. Pushing the space bar causes a "bullet" (it looks like a black dot) to appear and the cursor will move to the next line. Try inserting 10 to 20 bullets to start, each time you hear a natural break, by pushing the space bar, then you can push any key to stop the process, OR
- When you have come to the end of the recording, type @End.
- After you have done this, each time you place the cursor by a "bullet" and press F4, you will hear the segment of the recording linked to that line. Transcribe what you hear using the instructions found in this manual. In the section below, we will show you how to make this part even easier!

When you're finished linking the file, the transcript will look like this:



This looks very strange but is the beginning of something wonderful.

Next, go to the WINDOW menu and open the Walker Controller from the drop down menu. It looks like this:



Many of the terms in this box may seem like gibberish, but the ones you care about are: *walk length, loop number, and playback speed*. *Walk length* controls how much you hear when playing back a bullet (if you don't hear everything you should when you play a bullet, your segment is longer than the default, so just increase this number until you seem to hear enough). The *loop number* is how many times the segment will repeat while you type – very critical – pick the one that makes you happy! *Playback speed* controls how fast the playback occurs. 100% is normal, but if the interaction is going too fast, you can decrease rate (and pretend you are transcribing Darth Vader). Alternatively, if you are falling asleep because the person talks too slowly, you can actually make it go faster by setting playback speed to numbers greater than 100% (and transcribe Munchkins ☺). Keep the Walker Controller window open while you do the next steps below.

Click on the first bulleted line. Then click F6. ***(If this doesn't work in Mac**, go to System Preferences, Keyboard, and check "Use all F1, F2 keys as standard function keys.")* You should hear your segment repeat as many times as you told it to. You are now ready to transcribe. Here is another cool thing: unlike most programs, you can type while having the Walker open – typing won't deactivate the Walker window.

Now skip down to how to transcribe. Unless you are a perfectionist, in which case read the remainder of this section. Don't worry if your "bullets" (that's what the dots are called) don't line up exactly with utterances. THEY ARE PRIMARILY THERE JUST TO MAKE TRANSCRIBING EASIER AND MORE ACCURATE. NO PROGRAMS ACTUALLY COMPARE THE AUDIO/VIDEO BULLETS TO THE TRANSCRIPT. However, if you are very perfectionistic, you can easily change how much of an utterance is linked to which bullets. To do this:

- Put your cursor next to the bullet you want to change. (If you want to add more to the beginning, place the cursor next to the bullet above the one you want to change)
- Press F5. The audio will begin to play where that bullet starts
- Press the space bar when you want to end that bullet and begin the next one
- A black highlighting line will show you which line is currently recording
- When you've finished fixing the bullet(s), click the mouse anywhere and you will immediately stop recording bullets

HOW TO TRANSCRIBE

There are rules for this, for a reason. The programs expect transcripts to look a certain way, or they won't run, or they will generate gibberish, which makes the whole exercise meaningless. Once you get the hang of it, it's no different from any other activity in which you are asked to follow a format, such as a web address or a telephone number.[3]

Transcribing on Main Lines

Main Lines begin with a * and indicate what was actually said. A fast way to type in each speaker's 3-letter code is to pull down the Tiers menu on the top of your file. On the bottom you will see an option called "update". If you do this, the next time you pick the tiers tab, the first lines going down from the top will be the participants you've already identified in the header. Place your cursor in between the asterisk and the colon, press tiers, and either click on the appropriate one, or use the "hot key" to its right. You will get the speaker info inserted automatically.

Important: Each main line should contain only one C-unit.

WHAT THE HECK IS A C-UNIT? A conversational unit (in writing, we use the term T-unit, which is basically the same). A good rule of thumb follows Meatloaf's wonderful song, "Two out of three ain't bad".[4] Although defining an utterance (C-unit) may seem easy, it's a very real area of disagreement in transcription, as Stockman (2010)[5] notes [see these and other references in footnote below]. Because she notes increasing reliability with multiple cues, we have frequently used a "2 out of 3" criterion to define utterances in transcribing. The 3 features are:

[3] There is a shortened transcription manual for the AphasiaBank project at the TalkBank site if you wind up lost here; perhaps their approach will augment what you learn from this manual.

[4] If you haven't heard it, take a detour and visit:
http://www.youtube.com/watch?v=q8JA9Qs2Mho
Then, take a break and listen to "Paradise by the Dashboard Light", "I would do anything for love (but not that)", "You took the words right out of my mouth" and basically anything on the album, "Bat out of Hell" ☺

[5] Stockman, I. J. (2010). Listener reliability in assigning utterance boundaries in children's spontaneous speech. *Applied Psycholinguistics*, *31*(3), 363-395.

1. Silence or pause of more than 2 seconds
2. Terminal intonation contour
3. Syntax that makes a complete sentence, or word(s) that make a complete, appropriate contribution in conversation, as in, MOT: where are you going? CHI: home. (one word, but an utterance)

Our rule is: If you hear two or three of these, you have an utterance. If you hear only one, keep what follows as the same line, utterance, c-unit.

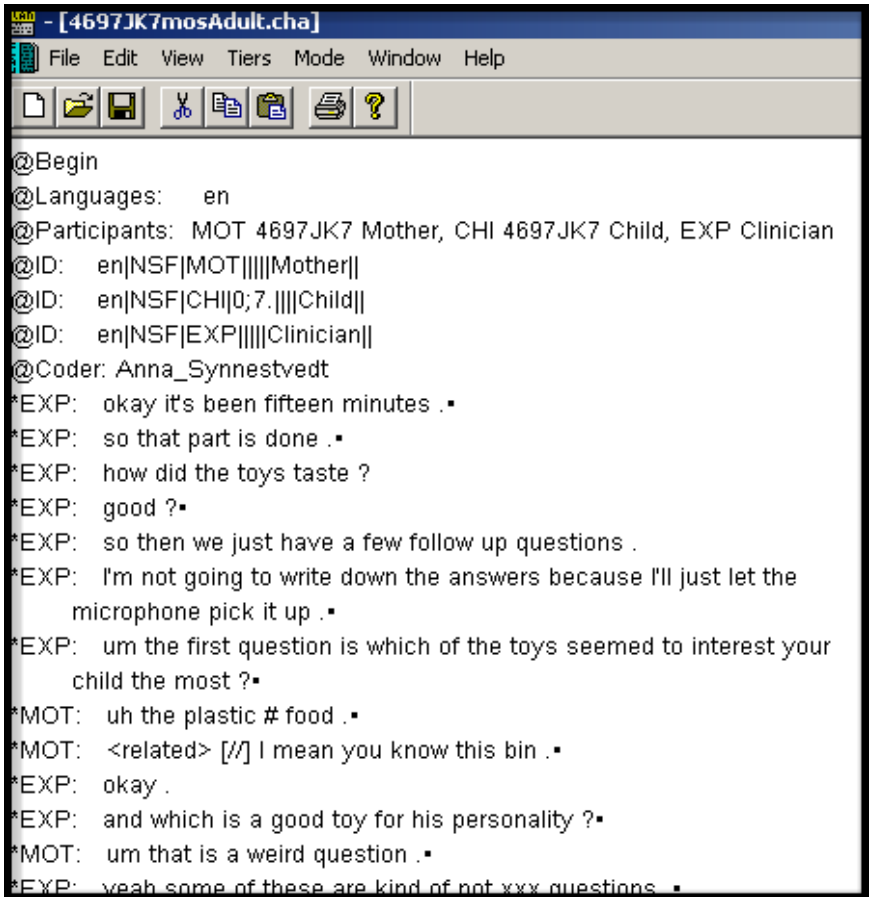
This makes it seem simple. In reality, this is the most subjective and unreliable part of transcription, which is sort of problematic, since so many analyses you want to run, such as MLU, require you to make good decisions about where utterances start and end. If you are doing research using transcripts, it is really important to compute reliability just for this reason alone, as Stockman found. (CLAN offers the RELY program to do this). For child language transcription, both Owens (2010, p 141-2) and Paul & Norbury (2012, p 303)[6] have handy charts to use in defining utterances. Most of their longer list reduces to the three criteria listed above. When a speaker produces several C-units in a row, code each with a new main line.

After the 3-letter code is a colon and then tab (remember, not spaces!!). As we noted above, when you linked the sound file to the transcript, CLAN automatically put in an asterisk and a colon followed by a tab and then the bullet. Using the Tier command then inserts the speaker code, or, if you are a Luddite (you can look up who they were), you can simply type the 3-letter code of the speaker between the asterisk and the colon and then move your cursor directly in front of the bullet and begin typing. This should ensure that the structure of the transcript is correct and consistent.

Type what the speaker says and end every line with some sort of punctuation, usually a period. Be sure to place a space between the end of the last word on a line and the punctuation mark.

Here is what a transcript looks like:

[6] Owens, R. (2010). *Language disorders* (5th ed). Boston: Allyn & Bacon; Paul, R. & Norbury, C. (2012).*Language disorders from infancy through adolescence*. Philadelphia, PA: Mosby.



Keep the transcript clean (one C-unit per line), even if the bulleted audio does not align exactly. Not every line must contain a bullet, but there is a *limit of one bullet per line.* Bullets may only occur at the end of the text after the punctuation.

Punctuation

There must be a delimiter (punctuation mark) at the end of every utterance. Most utterances will end in a period (.) question mark (?) or exclamation point (!). Be sure to put a space between the end of the last word and the punctuation mark.

No capital letters should be used except for "I" and proper nouns. There should be no punctuation inside of utterances (e.g., commas, semicolons, etc.) except for apostrophes. Some utterances will require different delimiters. These are described below.

Trailing Off +...

The trailing off or incompletion marker is the terminator for an incomplete, but not interrupted, utterance. Trailing off occurs when speakers shift attention away from what they are saying, sometimes even forgetting what they were going to say.

*CHI: smells good enough for +...
*MOT: what were you saying ?

If the speaker does not really get a chance to trail off before being interrupted by another speaker, then use the interruption marker +/. rather than the incompletion symbol. Do not use the incompletion marker to indicate simple pausing, repetition, or retracing.

Trailing Off of a Question +..?
If the utterance that is being trailed off has the shape of a question, then this symbol should be used.

*EXP: so do you have any of these toys at home or +..?

Interruption +/.
This symbol is used for an utterance that is incomplete because one speaker is interrupted by another speaker. Here is an example of an interruption:

*EXP: what did you +/.
*CHI: mommy.
*EXP: +, with your spoon.

Self Interruption +//.
This symbol is used for an utterance that is incomplete because one speaker interrupted him- or herself. Here is an example of a self-interruption:

*MOT: well we haven't started to +//.
*MOT: Alex put that down !

Retracing Without Correction [/] (also in fluency codes, Appendix 3)
The [/] symbol is used in those cases when a speaker begins to say something, stops and then repeats the earlier material without change. The material being retraced is enclosed in angle brackets. In a retracing without correction, it is necessarily the case that the material in angle brackets is the same as the material immediately following the [/] symbol. Here is an example of this:

*CHI: <I wanted> [/] I wanted to invite Margie .

If only one word is repeated, the angle brackets are not necessary. When the angle brackets are not used before the repeating symbol [/], CLAN assumes that only the last word was repeated.

If there are pauses (longer than 1 second) and fillers between the initial material and the retracing, they should be placed after the retracing symbol, as in:

*CHI: it's [/] (.) um (.) it's [/] it's like (.) a um (.) dog .

When a word or group of words is repeated several times with no fillers, all of the repetitions except for the last are placed into a single retracing.

Retracing with Correction [//] (also in fluency codes (Appendix 3))

This symbol is used when a speaker starts to say something, stops, repeats the basic phrase, but changes any part of the phrase. Usually, the correction moves closer to the standard form, but sometimes it moves away from it. The material being retraced is enclosed in angle brackets. In retracing with correction, it is necessarily true that the material in the angle brackets is different from what follows the retracing symbol. Here is an example of this:

*CHI: <I wanted> [//] uh I thought I wanted to invite Margie .

Retracing with correction can combine with retracing without correction, as in this example:

*CHI: <the fish is> [//] the [/] the fish are swimming.

Sometimes retracings can become quite complex and lengthy. This is particularly true in speakers with fluency and language disorders. It is important not to underestimate the extent to which retracing goes on in such transcripts.

When most programs run, information associated with repetitions and corrections is EXCLUDED; if you run FREQ, for instance, to see its options, it will tell you this; if you want this info to be counted in analyses, there is an option in the Command line (type in +r6 along with your other command information).

CLAN allows you to code for anything you want. See CLAN manual if you want to code for other aspects. Once you learn how to code a few things, you will be able to fiddle around with it.

Overlaps [>] [<]
Sometimes two people will speak at the same time. In most cases, you don't care, and sometimes, for clinical interactions, you don't really want to code one person just saying, "uh-huh" over and over again. Or, you can put the other person on the next line, even if they were talking at the same time. But if you care, here's how to code the overlap.

The "overlap follows" symbol ([>]) should be placed on the first line and the following line should have the "overlap precedes" symbol ([<]). See the example below.

*MOT: that <she can> [>] .
*EXP: <that she can stand on> [<] .

FLUENCY codes: *typical disfluency* (please see section on stuttering codes in Appendix 3 for specific codes and programs to transcribe and analyze stuttering)

Unfilled pauses
You can indicate short, medium and long pauses by using (.) (..) and (…), respectively.

Filled pauses
When words are used as fillers, they should be marked with &- to show that they should not be included in lexical analyses (i.e. MLU or TTR analyses) *and SHOULD be counted in disfluency analyses*. Examples:

&-like
&-you_know (note that we pulled the you and know together with an underscore so it gets counted as one lexical item)
&-and, when it seems to be just a staller…

NON-Lexical fillers/disfluencies (um, uh, er, etc.) should be transcribed as &-um, &-uh, etc. There is a fine line between interjections, communicative turn sounds, and these. Use the ampersand-dash combo (&-) to indicate what you think is a disfluency, as opposed to something that has a communicative function.

Quotation on Next Line +"/.
During story reading and similar activities, a great deal of talk may involve direct quotation. In order to mark off this material as quoted, a special symbol can be used, as in the following example:

*CHI: and then the little bear said +"/.
*CHI: +" please give me all of your honey.
*CHI: +" if you do, I'll carry you on my back.

The use of the +"/. symbol is linked to the use of the +" symbol. Breaking up quoted material in this way allows us to maintain the rule that each separate utterance should be on a separate line. You should only do this if the client is reading or quoting. If the child makes up utterances for a character, it's not all that important to mark it as quoted. You can just type what you hear. But some clinical and research questions like to notice when children or adults use another person's voice, because that is more sophisticated than simply narrating.

Quotation Precedes +"
This symbol is used when the material being directly quoted precedes the main clause, as in the following example:

*CHI: +" please give me all of your honey.
*CHI: the little bear said +".

Single quoted words:
Can use the @q marker, as in:
*CHI: and the boy said shh@q .

Multiple words that should hang together
Some words need to be treated as ONE word (e.g., "You know", "merry go round", "patty cake", or a person's name, such as Mister Spock, or Brian MacWhinney). When words link together to form a frozen phrase, link them with an underscore (you_know, patty_cake, Nan_Bernstein_Ratner). There are also complicated rules for using + to link generative compounds. If you get ? on a mor parse, this could be the reason, and you can play with typing it differently (pulling it together, pull it apart with underscore or +) to see if the program is happier. There shouldn't be too much of this in most files.

Other Coding Conventions: ERRORS!

Most of the time, people don't speak in perfect sentences. Here are some coding conventions for some of the most common deviations.

If you are reading this, you are probably a clinician ☺ If you wind up with a file that doesn't have some errors, we'd be amazed. That's WHY you do language sampling! When you find an error in use, mark it in the following way:

*CHI: he had two mouses [*] [: mice] .

As in your old linguistics class, an asterisk flags something ungrammatical. You can just use the [*] or you can code errors more finely (the CHAT and AphasiaBank manual has lots of examples), but the main point here is that you noted an error (which will be computed in KidEval and Eval), take away a DSS sentence point, and let the MOR program know what the person was trying to say.

If you know where the error occurred to make the sentence bad, you should put it after the mistake, or indicate where the missing word was (see 0 below). If the whole thing seems bad, but you can't identify a specific thing to flag, just put the [*] after the utterance delimiter, as what is called a post-code.

Common error codes after the asterisk:
semantic/lexical: * s
neologisms: * n
morphological: * m (there are many specific morphological codes if you want them; consult the CHAT manual error section)
Phonological errors: [please read the footnote]!.

Missing words: 0
Example 0does for

*CHI: 0does he like it? (this corresponds to the "GLOSS" (what you heard: He like it?)

Phonological Fragments &fr
If there is a fragment of a word use the & symbol and then type out the fragment, for example

*CHI: he had a &fr friend.
*CHI: I really wanted to &vi visit the zoo.

In the first example, the child began to say "friend" but stopped part way through the word. Notice that a lot of disfluencies are being tagged with the ampersand. The ampersand shows up in the Appendix under stuttering codes as well, for obvious reasons.

Unintelligible words xxx
If you cannot understand a word or phrase contained within an utterance, code it as xxx. Before doing so, try to listen a couple times to figure it out. If you think you can make out the sounds, you can use yyy, and then use a %pho tier to put these sounds in, but that's beyond the scope of this manual.

Pauses (.)
If the speaker pauses for a short duration, use the (.) symbol. This symbol is not necessary if the pause occurs between utterances, even if the bullet makes it sound like it happens at the beginning of the utterance. Pauses that occur between utterances do not need to be coded.

Sound Effects or Silence [=!]
This is rarely of interest unless you are working with a poorly verbal person for most clinical purposes. Sometimes, there aren't any words, but there is some activity or sound that is informative for understanding the transcript or the interaction. You mark it like this:

*MOT: you're just the cutest [=! kissing noises] .

If the only thing on the line is the sound effect, it should be coded like this:

*MOT: 0 [=! kissing noises] .

This format should also be used for bullets that only include sounds of toys moving around:

*MOT: 0 [=! toys rustling] .

This is also the format used when there are long periods of silence:

*MOT: 0 [=! silence] .

Notice that you have to actually note that 0 (nothing) was said, but that there was a conversational turn with non-verbal content. To note that nothing was said, use the number zero, not the uppercase letter 'O'.

Babbling and Jargon (from kids or patients)
Sometimes you get sounds, but you have no English word to "gloss". You can import fonts and insert phonetic symbols (or characters from other languages, such as Chinese) See the CHILDES website for help with this if you need it. For many situations, it may be best to mark it as unintelligible using yyy (and trying to gloss what you heard) or xxx (no real best guess as to how to transcribe):

*CHI: yyy [=! dada].

Reminder: Sounds or utterances that are truly unintelligible should be marked in this format:

*CHI: xxx [=! vocalizes/laughs/whines, etc] .

Now we have a bunch of Tables for various kinds of words. WHY? Because so many analyses will compare word types and tokens, having a system (and checking your spelling) is extremely important, or you will wind up crediting the speakers with more words than they really say. Sloppiness hurts your end result. A lot.

Exclamations, Interactional Markers and Other Funny Words - uh, um, sh

In order to keep spellings consistent, use these spellings for various non standard words:

Table 5: Exclamations

Exclamation	Meaning	Exclamation	Meaning
*ah	relief, joy	*pst	listen here
*ahhah	discovery	Sh	Silence
aw	sympathy	*tsk	Shame

golly	gee whiz	tut	pity
gosh	gee whiz	ugh	disgust, effort
ha(h)	Triumph	*uhoh	trouble
*haha	amusement	vroom	car noise
*heehee	amusement	whee	exuberance
*mmm	tasty, good	wow	amazement
*num	tasty	yea	a cheer
*nummy	tasty	(y)eek	fear
*numnum	tasty	y(o)ikes	mild fear
ouch	sudden pain	*yum	tasty
ow	hurt	yummy	tasty
oy	dismay	yumyum	tasty

Table 6: Communicators

Marker	Function	Marker	Function
ahem	ready to speak	nah	no
*emem	I don't know	uhhuh	yes
*er	pause	*uhhum	yes indeed
*hunmmm	no	*uhuh	no
*hunhunh	no	*uh	pause (any vowel)
huh	questioning	um	pause
hmm	thinking, waiting	ye(a)h	yes
hmm?	questioning	*yeahhuh	yes (contradicting)
*mmhm	yes	yep	yes
nope	no	yup	yes
*nuhuh	strong no	whoops	blunder

* Words that are marked with an asterisk cannot be found in Webster's Third New International Dictionary. The *** should not be included in the text of the transcript**.

There are actually many of these words in the lexicon file used by MOR; you can open up the co.cut file in lib, or any other grammatical file in Eng MOR to see what is in the "dictionary". But it might be just as easy to print out some of these CHAT manual tables and keep them around.

Shortened Words

Sometimes the word will be shortened but still intelligible. You can put the deleted part of the word in parentheses. This will enable CLAN to do analyses on the word, and will help you later if you are trying to count syllables, or need to know exactly how the child said the word. The following table lists some examples.

Table 3: Shortenings

Examples of Shortenings			
(a)bout	don('t)	(h)is	(re)frigerator
an(d)	(e)nough	(h)isself	(re)member
(a)n(d)	(e)spress(o)	-in(g)	sec(ond)
(a)fraid	(e)spresso	nothin(g)	s(up)pose
(a)gain	(es)presso	(i)n	(th)e
(a)nother	(ex)cept	(in)stead	(th)em
(a)round	(ex)cuse	Jag(uar)	(th)emselves
ave(nue)	(ex)cused	lib(r)ary	(th)ere
(a)way	(e)xcuse	Mass(achusetts)	(th)ese
(be)cause	(e)xcused	micro(phone)	(th)ey
(be)fore	(h)e	(pa)jamas	(to)gether
(be)hind	(h)er	(o)k	(to)mato
b(e)long	(h)ere	o(v)er	(to)morrow
b(e)longs	(h)erself	(po)tato	(to)night
Cad(illac)	(h)im	prob(ab)ly	(un)til
doc(tor)	(h)imself	(re)corder	wan(t)

Assimilations

Sometimes two or more words will blend together and sound completely different from the individual words that created them. In such cases you can type these assimilations into the main tier to reflect this change:

Table 4: Assimilations

Nonstandard	Standard	Nonstandard	Standard
coulda(ve)	could have	mighta	might have
dunno	don't know	need(t)a	need to
dyou	do you	oughta	ought to
gimme	give me	posta	supposed to
gonna	going to	shoulda(ve)	should have
gotta	got to	sorta	sort of
hadta	had to	sorta	sort of
hasta	has to	wanna	want to
hafta	have to	wassup	what's up
kinda	kind of	whaddya	what did you
lemme	let me	whyntcha	why didn't you
lotsa	lots of		

We've added: gotchya (ns) got you (s)

Transcribing Dependent Tiers

Why would you want to add tiers to your transcript? Well, common reasons are to mark for speech act intents, comments about what is going on, whether the client managed to include a narrative component in his story, etc. Dependent tier lines begin with the % symbol. This symbol is followed by a three-letter code in lowercase letters for the dependent tier type, such as "nar" for narrative code or "com" for a comment. The three letter code is followed by a colon and then a tab. The text of the dependent tier begins after the tab. A full list of pre-arranged, conventional tiers is in the CHAT manual. Consult it before making one up!

Comments %com

If you wish to note something odd or of interest in the recording, create a dependent tier for comments (%com).

*CHI: Alpha had a frog .
*CHI: then when Asa wake up he <could> [//] looked for his frog everywhere .
%com: the character's name changes from Alpha to Asa and back again .

Just like with main lines, there should always be a colon after the three-letter code and a tab after the colon before the actual comment begins. The same rules of punctuation and capital letters apply to these lines.

Why would you do this? It often helps to annotate something you noticed during the session that you are likely to forget later, that provides useful information about the utterance or interaction.

Note: If a significant portion of the file is unanalyzable, add a warning header at the top of the file, rather than comments all over the place; likewise, if there is a generic observation that affects the file, put it up at the top as well.

.

@Warning: The first 15 minutes of this file had background noise.

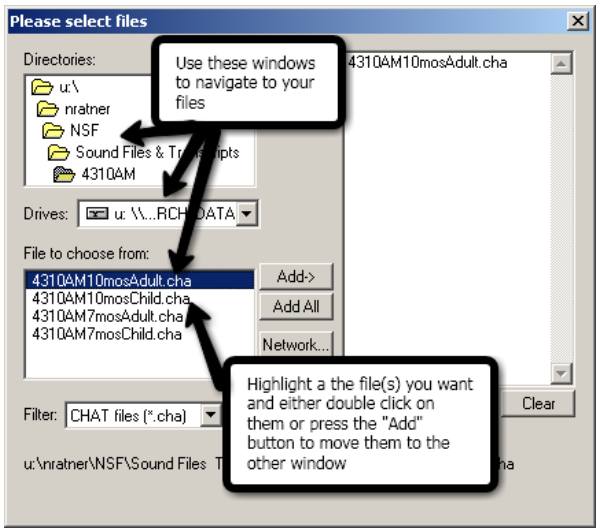
MOVING ON TO THE GOOD STUFF: CLAN ANALYSES

CHECK

Before you run any real analyses, it is important to make sure the transcripts are set up correctly or else the analyses won't run. To do this, you need to run a program called CHECK on the transcripts.

To run CHECK, open CLAN. The Commands window should automatically open. If it does not, press CTRL+D, or go to Window → Commands. In the commands window, type "check" followed by a space. Once you type in a command in the Commands window, a 'file in' button will appear above the window.

To choose the transcripts on which you'd like to run the analysis, click the "file in" button. A new window will open, and from there you navigate to where your transcripts are saved.



You can run CHECK and any of the following analyses on more than one transcript at a time if you so choose. If you do this, each file will be run separately in sequence. Once you have selected the file(s) you want to analyze and have moved them into the window on the right, press "Done." This will return you to the Commands window, and an "at" sign (@) will appear next to your command. Then press "Run" in the Commands window or simply hit Enter on your keyboard and the analysis will run on its own.

Results of most CLAN analyses are sent to a FILE OUTPUT that pops up on the computer screen (except for a few analyses we will explain further). If you can't remember all the mistakes, you can save the output that showed up on your screen to a file using the usual File command on the toolbar, or you can cut and paste the output to a word file and save it.

Typical CHECK gripes:

Wrong format (spaces, not tabs). Confusing spaces and tabs between the Speaker ID and the line of dialogue is such a common mistake, there is actually a program to fix

this. In the command window, type CHSTRING +q and then input the offending file. It will fix all the bad lines at once.

Missing headers. typically, this problem occurs when you didn't format appropriately, or didn't include a speaker you then transcribed, or invented a role not in the approved list.

Missing "utterance delimiters" (remember them? Periods, etc.). Sometimes it's easy to confuse a bullet for a period and forget to put some in. If you get a lot of these errors on a check run, there is a program to fix the situation: In the CLAN window, type DELIM and put in the offending file. DELIM will insert periods (but only periods) at the ends of utterances lacking final delimiters.

Forgetting to put @End at the end of the file ☹

Unrecognized symbols, people, etc. Typically, CHECK will actually underline and/or note the error in red what it doesn't like, and tell you the line it found it on, to help you go fix it (if it finds more than one error on a line, it will treat it as a completely different error, which can confuse, until you notice it had multiple sad things to say about line 61 or something like that). **COOL TIME AND LABOR-SAVING FACT: You can triple-click on the line number noted in the output and it will take you directly to that line or the error to fix it!**

CLAN actually runs two CHECKs when you run the CHECK command. Be sure to keep fixing the files until you get the "SUCCESS Second pass complete" message.

****HOWEVER, you <u>can</u> actually run most programs on files that CHECK is having some problems with; it may affect accuracy of output somewhat, but probably not in a clinically relevant way.** If a program won't run, it's because something is pretty wrong with your file, and you need to figure it out and fix it. *If you still have trouble, either contact Chi-Bolts (see listserv instructions) or e-mail one of us.*

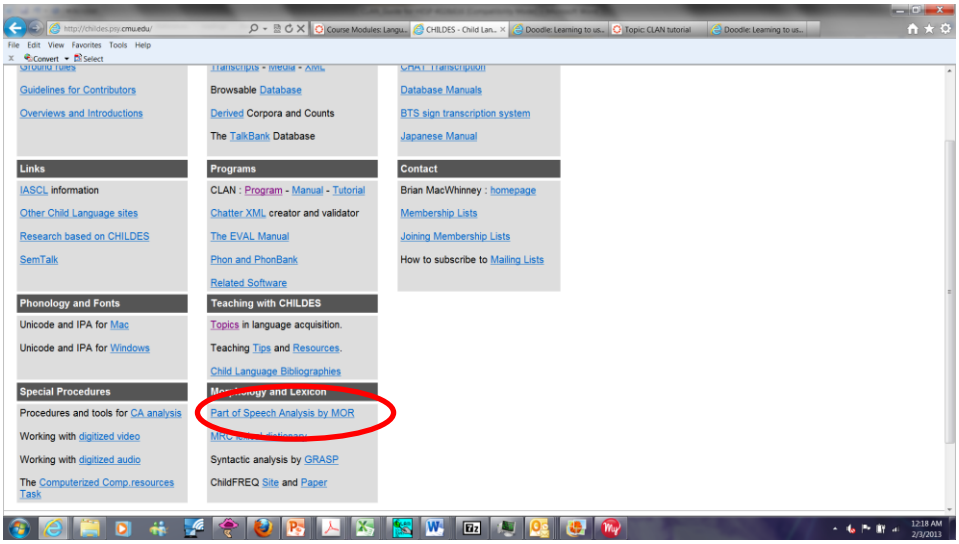
CLINICAL ANALYSES: KIDEVAL

Recently (Fall 2012), CLAN developed a super-useful utility called KIDEVAL (there is also EVAL for adult aphasia samples). This program sends its output automatically to an Excel spreadsheet with a mother lode of analysis results, such as mean length of utterance in morphemes (MLU-M), mean length of utterance in words (MLU-W), type-token ratio (TTR), clause density, and number of major morphemes observed. It will even output a series of transcript analyses into the same Excel file.

There is only one minor thing you have to do first to use KIDEVAL. You have to run a command called MOR on the transcript. MOR is short for "morphology" and will automatically insert a %mor tier, so that you don't have to figure out where morpheme boundaries are, or the syntactic construction of the sentence (e.g., parts of speech). Don't worry, your linguistics classes will still come in handy, but this saves a lot of work, and is much more accurate than doing stuff by hand.

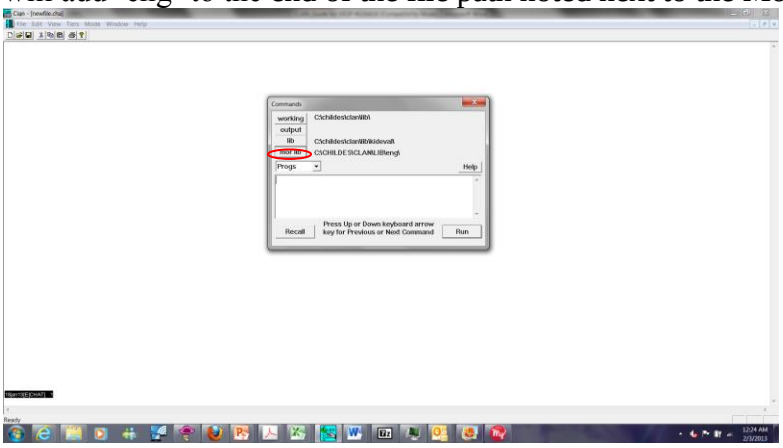
Preparing to run MOR

For MOR to run successfully, you need to configure your version of CLAN to include the very sophisticated and bulky rules that apply to each language (CLAN doesn't assume you are examining English – it is used all over the world.) You need to go back to the CHILDES project web site, and download the English MOR files (or the files for whatever language you put in the header and transcribed in). You will want to create a specific library directory for MOR that is distinct from the general CLAN lib directory. It is often convenient to place this MOR library inside the CLAN work or lib directory.



This will take you to the screen where you can download English (and/or any other language for which rules have been written; you can even analyze bilingual samples ☺; however, if you are going to analyze a single file with more than one language, that is an advanced topic, and you should contact Shelley Brundage [brundage@gwu.edu]).

Once you have downloaded the MOR library for English, you want to go back to the Commands window to do one simple thing: specify the MOR library you want CLAN to use for the upcoming analysis (in our labs, we often have people speaking in many languages, and we have to check to make sure we are not trying to create a morphological analysis of a Spanish speaker using English rules.) In the Command Window, click on the morlib button, then browse your directory to find the English MOR folder, and click to choose it. Doing this will add 'eng' to the end of the file path noted next to the Morlib button.



OK, you are ready to roll. In the CLAN commands window, just type MOR and select the file you want to analyze. IF YOU ADD THE STRING +1 TO ANY COMMAND, YOUR OUTPUT FILE HAS THE SAME NAME AS YOUR ORIGINAL FILE. That means that if I type: MOR +1 filename.cha, I will get a message that the program ran and produced a file with exactly the same name. That means that your original file has been *overwritten and whatever you asked it to do has been added to it*. So, be mindful of the +1 command. To add the MOR dependent tier, type:

Mor +1 filename.cha[7]

If you don't use +1 (which is a standard instruction to rewrite the same file you are working on), it will just produce a new file that starts with the same filename, but ends in .mor.cex. This is not a problem, but can create a lot of files if you are handling more than a single case at a time. That's why we usually overwrite the old file by using +1. But if you are leery, make a backup copy on a flash drive in case you ruin the original file.

We have found that a common error in running MOR is that MOR gives you an error code that it couldn't find something like ar.cut, which is one of the first grammar files MOR wants to use. If you get this, take a look in your computer for the Eng folder. Sometimes it winds up somewhere not exactly where you thought you put it and told CLAN the MOR library was. In short, if you get this error, find the Eng folder using your computer utility, and get the locations to match, by either moving Eng, or changing where the mor library is supposed to be in CLAN. We have even seen Eng get embedded inside ANOTHER Eng folder, so click around if you get this message.

For ENGLISH, MOR automatically runs a second program called POST. If you are analyzing English, you are ready to run any CLAN program, including the KidEval/Eval suite.

The following section is for hard-core language people: MOR is actually now TWO programs, MOR and POST, for many languages (English, Spanish, French, German, Hebrew, Japanese, Mandarin and Cantonese, with more languages coming); IF YOU ARE USING ONE OF THESE LANGUAGES, AND THIS STUFF BORES YOU, SKIP to ***KIDEVAL****.*

IF YOU ARE USING ANOTHER LANGUAGE, PLEASE CONTINUE TO READ,

IF YOU ARE USING A LANGUAGE THAT HAS A MOR GRAMMAR BUT IS ***NOT*** *ENGLISH, SPANISH, FRENCH, GERMAN, HEBREW, JAPANESE, MANDARIN OR CANTONESE, YOU WILL NEED TO CLEAN UP YOUR OUTPUT BY HAND:*

[7] If you have a file with lots of non-English/mistakes in word use or grammar, you might want to run mor +xb <filename>; this will produce a file ending in .cex that has the words it couldn't find in its dictionary (e.g., something like "mouses"); this will tell you what line the odd word is on. Triple click on the file and it will open; triple click on the word, and it will take you to the transcript line. Add the right root or word right after the error, as in [: mice]. If you do this while transcribing, obviously, you are in better shape. If you find a lot of unrecognized words that you know ARE words, let the folks at CMU know (in any language that uses Mor). If you find only one or two, to be honest, your clinical measures probably won't be off by much. They'd probably be worse if you did it by hand, frankly.

No matter how smart the linguists are, a first pass with MOR produces some ambiguous parses (we will use English examples here, to illustrate the point):

*CHI: and they saw <the dog> [//] uh the frog was missing .
%mor: coord|and pro:sub|they v|saw^v|see&PAST^n|saw ?|uh det|the n|frog v:cop|be&PAST&13S^aux|be&PAST&13S adj:part|miss-PROG^n:gerund|miss-PROG^part|miss-PROG .

This is sort of messy, but the important concept is that some options are connected by "^" (section highlighted in blue in above example). This means that the program can't tell which option is right (in the example above, whether "saw" is the verb "to saw", or the past tense of "see" (which is the right answer) or the thing you cut wood with (the noun "saw").

In English, Spanish, Japanese, Mandarin, French, Hebrew, Cantonese, and German, a program called POST runs automatically to figure out the right answer and disambiguate. But for other languages, we don't have that utility yet, although more keep coming every week. So, you have to "clean up" your file by hand, which really doesn't take too long.

How do you do this? You just delete the wrong parse(s) and keep the right one by knowing the right answer (which is usually pretty easy for us language specialists). If your file is long, you can use the control-F option to find the "^" symbols. If you do find some, you need to go in, look at the line, and decide which answer you like. Then simply delete the WRONG string on the line by backspacing over it, and resave the file.

If you see a "?" on a word you think *is* a good word of a language, it's likely that you made a compound out of separate words, or vice-versa, or spelled incorrectly. For fast clinical work, you can replace the ? with the right part of speech (e.g., N for noun, V for verb, or any of the grammatical abbreviations you will see all over the mor tiers in the file), and all is well. (Good enough for most everyday work).

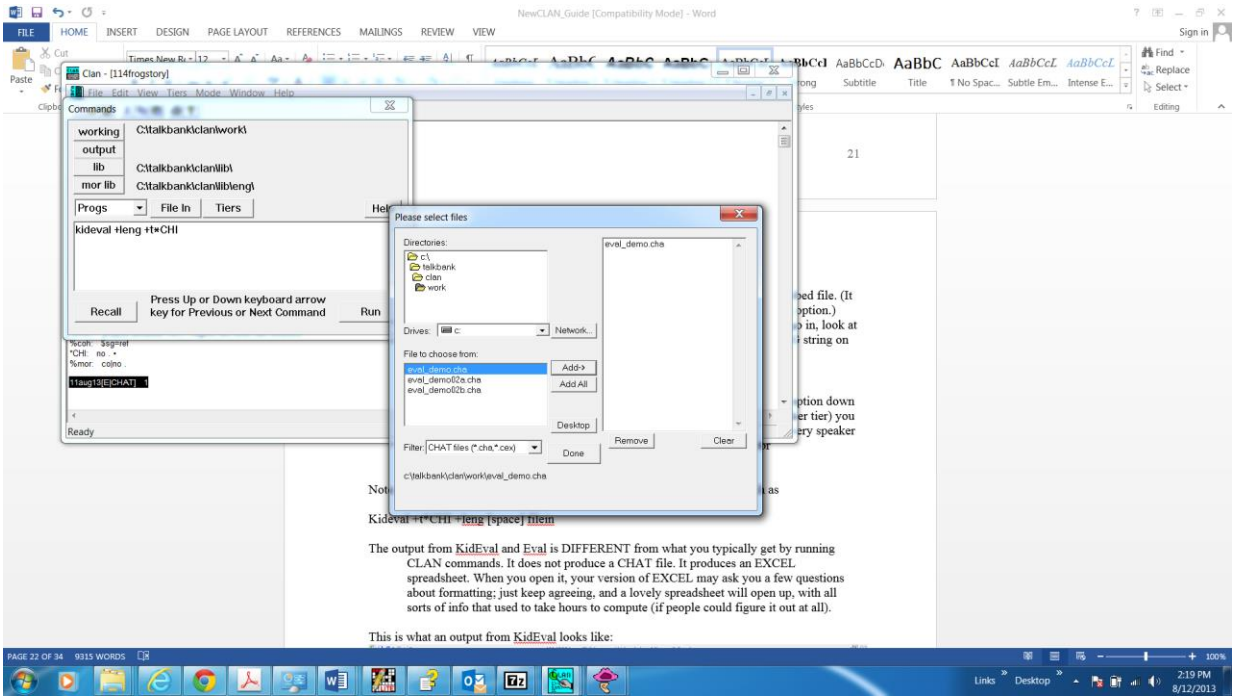
Then, it is good to CHECK the file again. If there's a problem, you really should go back and make a fix. This is typically that POST didn't disambiguate a line and there is still a ^ separating grammatical options.

KIDEVAL: the answer to the SLP's dreams for child (and adult) language sampling.
STAY TUNED for **EVAL**, below, which analyzes language samples from adults along dimensions more typically relevant for aphasia.

Once you have your file automatically MOR-ed and you checked to make sure there aren't lots of ? marks under perfectly good words, just type KIDEVAL in the CLAN Command window. It will ask you who (which speaker tier) you want to analyze, which makes sense, since you don't typically care about every speaker in the sample. You must also pick a language option (such as +leng for language=English) Click and you are done (VOILA!)

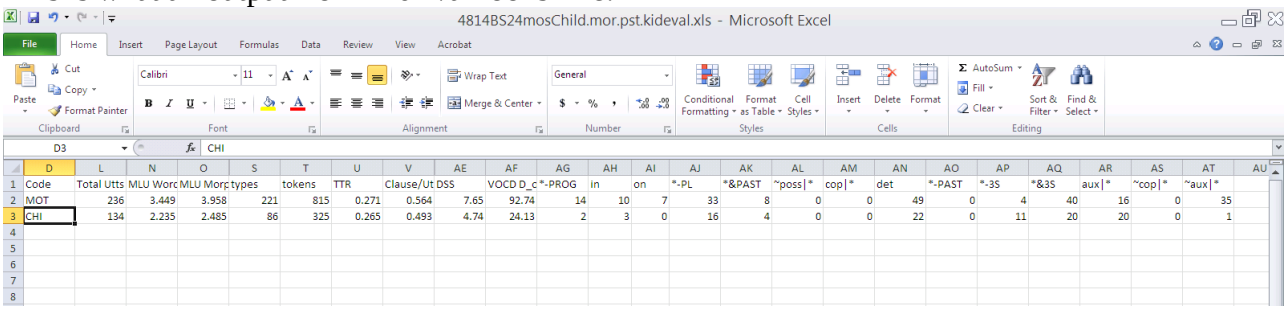
Note: you must put SPACES in between the parts of the command line, such as

> Kideval [space] +t*CHI [space] +leng [space] click file in [then pick the file/files you want]



The output from KidEval and Eval is DIFFERENT from what you typically get by running CLAN commands. It does not put the data in a CLAN output file. It produces an Excel spreadsheet that you open in Excel (the output file will tell you where CLAN 'put" the spreadsheet). When you open it, your version of Excel may ask you a few questions about formatting; just keep agreeing, and a lovely spreadsheet will open up, with all sorts of info that used to take hours to compute (if people could figure it out at all, or do the linguistics or math right).

This is what an output from KidEval looks like:



	D	L	N	O	S	T	U	V	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU
1	Code	Total Utts	MLU Word	MLU Morp	types	tokens	TTR	Clause/Ut	DSS	VOCD D_c	*-PROG	in	on	*-PL	*&PAST	~poss|*	cop|*	det	*-PAST	*-3S	*&3S	aux|*	~cop|*	~aux|*	
2	MOT	236	3.449	3.958	221	815	0.271	0.564	7.65	92.74	14	10	7	33	8	0	0	49	0	4	40	16	0	35	
3	CHI	134	2.235	2.485	86	325	0.265	0.493	4.74	24.13	2	3	0	16	4	0	0	22	0	11	20	20	0	1	

This is a big display, so here is a fast guide to what you are looking at (we have actually hidden some columns you would get if ran the program, just to fit the most clinically useful stuff into the screen shot.)

Each person's data goes on a single row (there was both a Mother and Child in this file, and we looked at both of them, just for fun). Going across the top, you see, in order:

- The total number of utterances the speaker produced

- Mean Length of Utterance (MLU) in words (this is strictly not MLU, which is measured in morphemes, but is very good for looking at other languages.)

- MLU in morphemes (the traditional measure)

- MLU and MLU-W have recently updated "norms" for children across childhood (2;6-8;11). See Rice et al., (2010) Mean Length of Utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments. Journal of Speech, Language, and Hearing Research, 53, 333–349. (See Appendix 1).

- Types: the number of different words used in the sample (there are options in CLAN for lumping word roots and inflections, but this one takes each word version separately)

- Tokens: the total number of all words in the sample

- TTR: the Type-Token ratio, a measure of lexical diversity. This number goes from .001 to 1.0; a low value indicates a lot of repetition (as if someone said "yada yada" 100 times and nothing else); a high value means each word in the sample was different. You actually can't get numbers too close to 1.0 and still have grammar – think about it.

- There are a number of references showing typical TTR values for children across early development. One (Miller et al., 2005) is included in Appendix 2.

- Clause/utt: clausal density, the average number of clauses per utterance.

- DSS: Developmental Sentence Score. This normed value is a classic in clinical child language sample analysis. The manual refers you to source material and norms (we have pdf files of the scoring procedures and norms if you'd like them; due to copyright, we provide them separately). REMINDER: DSS gives a sentence point to anything you didn't place an error code on. If you didn't mark utterances as ungrammatical when you saw them, this value will be somewhat higher than it should be when comparing norms.

- VocD: Vocabulary diversity. This is a twist on TTR, but supposedly more reliable when looking at files that vary substantially in length. The CLAN manual has lots of information about this measure. It is used by many researchers instead of TTR. *For fluency researchers, we have serious concerns about all other vocabulary diversity measures, because PWS tend to talk so much less in most sampling sessions.*

- *-PROG through ~aux|* : These columns provide "Brown's morphemes", in developmental order, as they were seen in the sample. These numbers are either RAW

COUNTS, or percentages, depending upon which *option* for KidEval you pick[8] and add to the command line. [9]

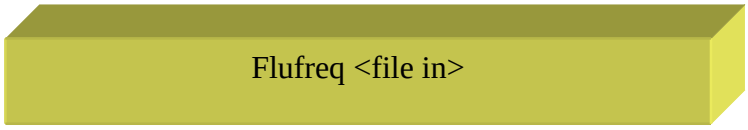
EVAL

EVAL is for clinicians who want to know about ADULTS' language abilities, and was developed to analyze the speech of persons with aphasia. When you type (or choose) eval in the command window, it will ask you to pick a file, then pick among options. If you like, it can compare your client to a database of people in AphasiaBank, a large cooperative endeavor funded by TalkBank (Brian MacWhinney and Audrey Holland, PIs). There is a full Eval tutorial up at the CHILDES website. Basically, if you can run KidEval, you can run Eval. It requires the same %mor line, requires that you pick +leng to make sure it runs using the English mor library and produces Excel output.

If, under options, you pick "Update database", and you pick controls for your comparison group, the resulting Excel spreadsheet will flag (with asterisks), categories in which your patient differed by more than a standard deviation from the control adults who were studied in the AphasiaBank initiative. We think this is pretty cool, frankly [actually, one reviewer of this manual used lots stronger "linguistic enthusiasm"].

FLUENCY ANALYSES IN CLAN:

You can run fluency AND language analyses on the same file. To produce a list of disfluent words, separated by type of disfluency, all you need to do is use the following command:



The output will look like this:

Speaker: *TOM:
6 &-like[10]
2 &-or_whatever
1 &-or_you_know
6 &-uh
7 &-um
1 &-you
14 &-you_know
7 [/] *(these are whole word repetitions)*
1 f:ailures
1 f:orward
1 s:ide

[8] To get raw counts for this program, add the command +o4.
[9] To see what kinds of options exist for ANY program, go to the command window and just pull down or type the program name, then click RUN. The screen output will give you the proper syntax and any unique options that exist for that program.
[10] Anything that starts with & is a filled pause

1 s:ometimes
1 s:tuttering
1 ↫a↫all
1 ↫h↫how
1 ↫i↫in
1 ↫ps↫ps:ychological
1 ↫th↫that's
1 ↫t↫teacher
1 ≠&-you_know
1 ≠being
1 ≠group
1 ≠present
1 ≠which
1 ≠worst
---
25 Total number of different item types used
61 Total number of items (tokens) *(number of SLD's to place over freq count of all words to get % stuttered words*

0.410 Type/Token ratio

YOU ARE NOW DONE WITH EVERYTHING YOU NEED TO KNOW ABOUT RUNNING CLINICAL ANALYSES USING CLAN.
YOU CAN STILL ASK OTHER CLINICAL AND/OR RESEARCH QUESTIONS USING THE SAME TRANSCRIPT, SO HERE ARE SOME MORE PROGRAMS TO LOOK AT.
YOU CAN ALSO GO TO THE CLAN MANUAL AND BE AMAZED AT HOW MANY THINGS YOU CAN DO WITH A SINGLE TRANSCRIPT.
HAVE FUN!

JOINING CHI-BOLTS

We predict you will really like CLAN after the rather steep initial learning curve; if you think you will use it after being MADE to use it in any class assignments, we suggest you do something pretty simple to stay up to date and figure out "glitches" or new ways to use the system. Go back up to the CHILDES Project site; under contacts, there is an option to join membership lists. CHI-Bolts is the non-intrusive but helpful listserv that helps people who get frustrated, who seem to have located problems (bugs), who wish Brian MacWhinney & Leonid Spektor would write a new program or option [sometimes, their wish is fulfilled within hours], or who want to lurk and see how people get creative. You won't get more than one note a week, typically.

If you think you have found a "bug", please do the following: send the list your sample file, the command string you used, and the nasty message CLAN sent you instead of the printout that you hoped for.

CONVERTING SALT FILES

Are you a SALT user? If we have won you over to the extent that you wish you could analyze OLD files using CHAT/CLAN, we can still help you. If you have a SALT file, the SALTIN command will turn the file to CHAT and then, typically with few fixes, you can then start from the point at which you run Mor, Post, etc. to get Eval or KidEval output to compare. If you have totally unformatted text, the TEXTIN command takes files, no matter how messy (one sentence on a line, or whole paragraphs), and generates a rough CHAT transcript that you can clean up and work with using the information in this guide.

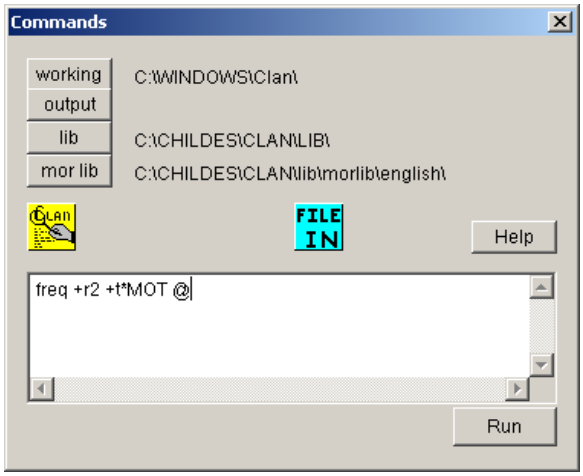
OTHER COMMANDS AND PROGRAMS PEOPLE USE A LOT:

For all the CLAN commands listed here if you type or select a clan utility and press 'run', you will get a printout that specifies the command line and the many options for 'tailoring' that program to your needs. There is also a CLAN "glossary" with some simplified suggestions at the AphasiaBank site on TalkBank.org

FREQ

FREQ is a command that can be run in CLAN once a transcript is completed that lists all of the words used in that transcript and a count of how many times each word is used. BECAUSE IT CAN BE CUSTOMIZED, FREQ IS THE MOST POPULAR AND USEFUL PROGRAM IN CLAN. It's normally used to compute the Type-Token Ratio (TTR), but can be used for much more.

To run FREQ for spell-checking, simply type "freq +r1" into the Commands window followed by a space. To run FREQ for target identification, type "freq +t*MOT" followed by a space into the Commands window, as shown above. "Freq" is the command. The t*MOT tells the program to only run on lines beginning with *MOT. When the analysis is run with no tier identification, it will automatically run on all main lines.



Once you have typed in your command, click the "file in" button and choose the file(s) on which you wish to run FREQ in the same way you chose files for CHECK. Then click Run on the Commands window or press Enter on your keyboard to run the analysis. An output file will open with a list of all words used in the transcripts.

If you are spell-checking, read through the words and identify any misspellings or miscodings. Common things to look for are: words connected to punctuation marks, which will happen if spaces are not inserted between the end of the utterance and the punctuation or if an error has been made in the coding; typing xx when it should be xxx for unintelligible words or phrases; and misspellings of words such as "spahgetti." (see KWAL below for how to fix this kind of nasty mistake).

The manual for CLAN shows the many things you can count using FREQ, which are virtually **unlimited** (which is why most researchers spend most of their lives using this single option). A widely used option is to make a text file with a class of words you'd like to track (such as fluency codes, question words, pronouns, etc.) and then FREQ your language sample using that file as input. That can be a way to check which pronouns the child has or doesn't yet have, etc. It works just like freq, except that you use the option +s@<filename> in the command line, where "+s" means 'search for' what is described in the file specified. So, if I made a file with question words, one per line, and saved it as questions.cut in the lib directory (we usually use cut to distinguish these files for actual language samples), we could write:

Freq +t*CHI +s@questions.cut <filename>

And the output would list how many times the child used what, who, where, when, why and how. Examples of how to use FREQ to search common kinds of things that clinicians or researchers might want are listed in great detail in the "real" CLAN manual.

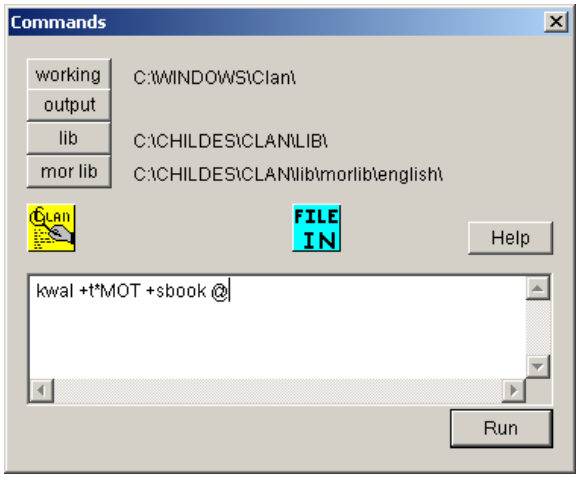
There is already a "fluency" folder inside CLAN/LIB that contains a utility file with all the stuttering codes. It is used to calculate fluency rates, and can be used for other analyses as well (if you are a researcher). For instance, you could insert the stutter codes from the main tier on

dependent tiers, such as the %mor line. Then you could calculate the rate of stuttering on various parts of speech[11].

KWAL (for KeyWordAndLine)

KWAL is another analysis run by CLAN using the Commands window. This analysis takes a word and finds the lines on which that word occurs in each transcript. This analysis is necessary to find out which lines the targets are on and in what position in the utterance each target is located.

To run KWAL, type "kwal +t*MOT +s" followed by the target of choice and then a space into the commands window. Be sure to type a space before each plus sign and after your target word.

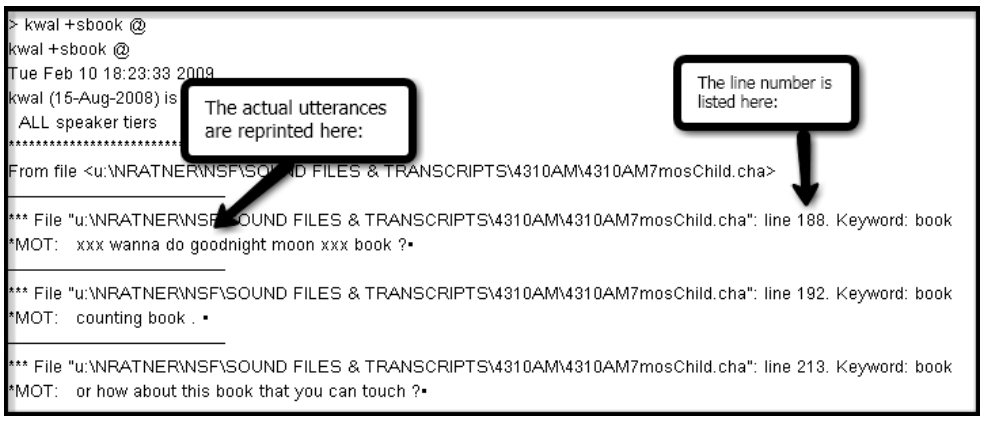


(If the target you are looking up contains an apostrophe, enclose the target in quotation marks. For example: kwal +t*MOT +s"he's")

Then click on the "file in" button and follow the same instructions for choosing your files as in the FREQ analysis. Click "Run" in the Commands window or press Enter on your keyboard to run the analysis.

The output will show an entry for each instance of that word used in the transcript and it will tell you on what lines the target occurred.

[11] For an example, see Tsai, Peitzu, Brundage, Shelley, Lim, Valerie and Bernstein Ratner, Nan (2010). Linguistic analysis of bilingual stuttering: concepts and methods. In P. Howell & J. Van Borsel (eds.). *Fluency disorders and language diversity*. Taylor & Francis.



If you triple click on the line that has the triple asterisks, it will take you right back to the line in the transcript. This is great for fixing mistakes if you find them at some point. KWAL the mistake, and the program will provide output that you can click to go right back the error and fix it.

COMBO

COMBO finds combinations of things that you might want to count or track; for example, if you wanted to know how often a child said "Once upon a time", you would have to use COMBO because this is a multi-word string. Likewise, if you wanted to see how often a child used the aux with a participle in the present progressive on the %mor tier, you would use a combo command, like this:

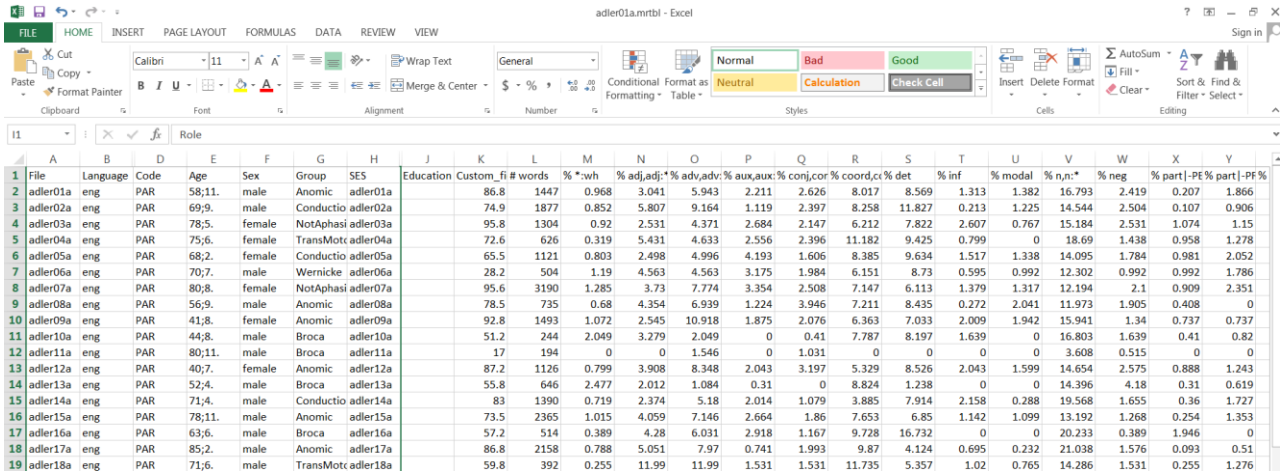
combo +t%mor +s"aux|*^part|*-PROG" *.cha

MORTABLE

Mortable will generate an excel spreadsheet that computes the proportion of particular syntactic elements (not raw counts), in a sample, quite an excellent clinical assessment or tracking ability. The command for mortable is:

mortable +t*CHI +leng <filein>

Like KidEval, this output will be an Excel spread sheet. Here is a sample output, from some aphasia data (it used +u to include all the files in a directory of clients)



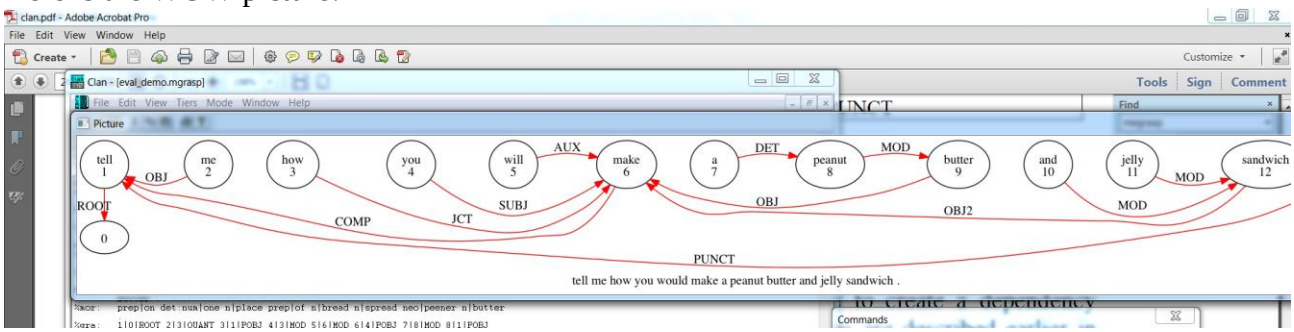
	A	B	D	E	F	G	H	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1	File	Language	Code	Age	Sex	Group	SES	Education	Custom_fi	# words	% *:wh	% adj,adj:*	% adv,adv:	% aux,aux:	% conj,cor	% coord,cc	% det	% inf	% modal	% n,n:*	% neg	% part|-PE	% part|-PF %
2	adler01a	eng	PAR	58;11.	male	Anomic	adler01a		86.8	1447	0.968	3.041	5.943	2.211	2.626	8.017	8.569	1.313	1.382	16.793	2.419	0.207	1.866
3	adler02a	eng	PAR	69;9.	male	Conductio	adler02a		74.9	1877	0.852	5.807	9.164	1.119	2.397	8.258	11.827	0.213	1.225	14.544	2.504	0.107	0.906
4	adler03a	eng	PAR	78;5.	female	NotAphasi	adler03a		95.8	1304	0.92	2.531	4.371	2.684	2.147	6.212	7.822	2.607	0.767	15.184	2.531	1.074	1.15
5	adler04a	eng	PAR	75;6.	female	TransMotc	adler04a		72.6	626	0.319	5.431	4.633	2.556	2.396	11.182	9.425	0.799	0	18.69	1.438	0.958	1.278
6	adler05a	eng	PAR	68;2.	female	Conductio	adler05a		65.5	1121	0.803	2.498	4.996	4.193	1.606	8.385	9.634	1.517	1.338	14.095	1.784	0.981	2.052
7	adler06a	eng	PAR	70;7.	male	Wernicke	adler06a		28.2	504	1.19	4.563	4.563	3.175	1.984	6.151	8.73	0.595	0.992	12.302	0.992	0.992	1.786
8	adler07a	eng	PAR	80;8.	female	NotAphasi	adler07a		95.6	3190	1.285	3.73	7.774	3.354	2.508	7.147	6.113	1.379	1.317	12.194	2.1	0.909	2.351
9	adler08a	eng	PAR	56;9.	male	Anomic	adler08a		78.5	735	0.68	4.354	6.939	1.224	3.946	7.211	8.435	0.272	2.041	11.973	1.905	0.408	0
10	adler09a	eng	PAR	41;8.	female	Anomic	adler09a		92.8	1493	1.072	2.545	10.918	1.875	2.076	6.363	7.033	2.009	1.942	15.941	1.34	0.737	0.737
11	adler10a	eng	PAR	44;8.	male	Broca	adler10a		51.2	244	2.049	3.279	2.049	0	0.41	7.787	8.197	1.639	0	16.803	1.639	0.41	0.82
12	adler11a	eng	PAR	80;11.	male	Broca	adler11a		17	194	0	0	1.546	0	1.031	0	0	0	0	3.608	0.515	0	0
13	adler12a	eng	PAR	40;7.	female	Anomic	adler12a		87.2	1126	0.799	3.908	8.348	2.043	3.197	5.329	8.526	2.043	1.599	14.654	2.575	0.888	1.243
14	adler13a	eng	PAR	52;4.	male	Broca	adler13a		55.8	646	2.477	2.012	1.084	0.31	0	8.824	1.238	0	0	14.396	4.18	0.31	0.619
15	adler14a	eng	PAR	71;4.	male	Conductio	adler14a		83	1390	0.719	2.374	5.18	2.014	1.079	3.885	7.914	2.158	0.288	19.568	1.655	0.36	1.727
16	adler15a	eng	PAR	78;11.	male	Anomic	adler15a		73.5	2365	1.015	4.059	7.146	2.664	1.86	7.653	6.85	1.142	1.099	13.192	1.268	0.254	1.353
17	adler16a	eng	PAR	63;6.	male	Broca	adler16a		57.2	514	0.389	4.28	6.031	2.918	1.167	9.728	16.732	0	0	20.233	0.389	1.946	0
18	adler17a	eng	PAR	85;2.	male	Anomic	adler17a		86.8	2158	0.788	5.051	7.97	0.741	1.993	9.87	4.124	0.695	0.232	21.038	1.576	0.093	0.51
19	adler18a	eng	PAR	71;6.	male	TransMotc	adler18a		59.8	392	0.255	11.99	11.99	1.531	1.531	11.735	5.357	1.02	0.765	14.286	1.531	0.255	1.276

If you run mortable on two different files, for instance your client's baseline file and then another file some sessions later, it would allow you to compare how much more (hopefully) your client uses a target morpheme under treatment.

MEGRASP

Currently, MEGRASP isn't used by any clinical programs, although it was developed for IPSYN, another powerful child language sampling procedure with norms that KidEval measures. But, if you always wanted to tree diagram something, if a transcript has been mor –ed, you can type MEGRASP then file in, and you will generate a file that also now has a %GRA tier under the %MOR tier. If you triple click on THAT, you will actually see a tree diagram pop up for your utterance. *This might be a fun party trick if a linguistics get-together needs something to liven it up. Or you want to help your kid with burdensome English homework* ☺We have included it here for its WOW value. Stay tuned for more developments. CLAN develops at the speed of light. And it never costs you anything to upgrade.

Here is the WOW picture:



More information for people who want to know more/too much:

OPTIONAL TRANSCRIPTION OPTIONS

Manually adjusting bullet length:

First, go to Mode menu and choose expand bullets. The control (start bullet) and alt keys (end bullet) can adjust bullet lengths up and down slightly to get full utterances on a bullet. Push alt or ctrl key and right or left arrow to go up or down by set intervals. This is an option as opposed to re-bulleting a finished transcript to get the bullets to align more perfectly. (Again,

try to suppress spending a lot of time to make them perfectly align because the bullets for most clinical work (and even research) is more to help speed transcription; it does not enter into most analyses.

Quick insert of speaker tiers:
When you are on an empty bulleted tier, go to "tiers", use the pull down menu to click on the speaker you want to insert and either click on it or use the hot key option that is listed next to that speaker.

"Lazy GEMS":
Use this command to mark different activities in one interaction. (see tutorial 14) Suppose you have different activities in one session, and you don't want KidEval or another program to include some of them. You can use lazy gems to mark the beginnings of different activities, e.g. @gem talking, @gem reading (which you don't want, because it is not spontaneous language), and run your programs using the option to do only "talking" gem sections.

WE THINK THIS IS MORE THAN ENOUGH TO GET YOU STARTED USING CHAT MEDIA-LINKED TRANSCRIPTION AND CLAN. WE HOPE YOU LIKE IT, WILL USE IT, AND WILL PROVIDE US WITH FEEDBACK. WE ARE ALSO HERE TO HELP. ENJOY![12]

[12] With thanks to the US NIH, for its support of CHILDES, Aphasia Bank and Phon and NSF for its support of TalkBank.

APPENDIX 1

Reference values for children's MLU and TTR.

MLUw and MLUm (From Rice et al., 2010)

Table 4. Means and standard deviations for mean length of utterance in words (MLU_w) and morphemes (MLU_m) per group.

		MLU_w			MLU_m		
Age range	*n*	*M*	*SD*	Cohen's *d* effect size	*M*	*SD*	Cohen's *d* effect size
				Affected			
2;6–2;11	6	2.37	0.32	0.93	2.59	0.39	0.90
3;0–3;5	15	2.84	0.38	0.97	3.07	0.48	1.07
3;6–3;11	24	3.10	0.75	1.04	3.36	0.80	1.09
4;0–4;5	54	3.31	0.70	1.22	3.64	0.80	1.22
4;6–4;11	72	3.60	0.62	0.95	3.95	0.70	1.01
5;0–5;5	84	3.72	0.61	1.05	4.09	0.70	1.10
5;6–5;11	97	3.95	0.60	0.85	4.34	0.67	0.89
6;0–6;5	108	3.98	0.70	0.90	4.38	0.75	0.92
6;6–6;11	94	4.18	0.71	0.79	4.63	0.79	0.84
7;0–7;5	103	4.15	0.62	0.69	4.56	0.69	0.72
7;6–7;11	100	4.33	0.88	0.57	4.77	0.96	0.60
8;0–8;5	94	4.36	0.75	0.86	4.80	0.83	0.89
8;6–8;11	61	4.49	0.86	0.72	4.97	0.93	0.68
				Unaffected			
2;6–2;11	17	2.91	0.58		3.23	0.71	
3;0–3;5	29	3.43	0.61		3.81	0.69	
3;6–3;11	38	3.71	0.58		4.09	0.67	
4;0–4;5	49	4.10	0.65		4.57	0.76	
4;6–4;11	74	4.28	0.72		4.75	0.79	
5;0–5;5	78	4.38	0.63		4.88	0.72	
5;6–5;11	77	4.47	0.61		4.96	0.70	
6;0–6;5	70	4.57	0.66		5.07	0.75	
6;6–6;11	63	4.70	0.66		5.22	0.71	
7;0–7;5	51	4.72	0.83		5.22	0.91	
7;6–7;11	47	4.92	1.03		5.45	1.13	
8;0–8;5	41	5.08	0.84		5.67	0.97	
8;6 – 8;11	18	4.99	0.71		5.51	0.79	

Type-Token Norms

Total Number Of Words And Different Words In T-Unit Sample Based On 50 Utterances					
Age	Words	Conversation		Narrative	
		Mean	SD	Mean	SD
3	Total	160.53	33.91	194.26	49.97
	Different	76.35	16.86	83.41	17.84
4	Total	195.78	60.68	257.19	48.17
	Different	92.07	24.41	108.26	17.75
5	Total	254.48	45.79	277.82	54.75
	Different	114.91	18.97	110.12	15.47
6	Total	251.83	59.93	325.31	80.25
	Different	112.95	20.53	121.44	21.11
7	Total	287.05	62.04	379.46	78.64
	Different	124.89	17.94	132.97	20.53
9	Total	311.70	66.61	419.41	73.43
	Different	133.48	21.48	133.26	25.11
11	Total	342.41	99.39	465.44	84.24
	Different	141.70	29.12	168.70	24.69
13	Total	330.15	78.92	447.96	72.02
	Different	139.00	24.92	159.11	21.17

Miller, J., Long, S., McKinley, N., Thormann, S., Jones, M.A., and Nockerts, A. (2005). *Language sample analysis II: The Wisconsin guide (pp. 21-22). Madison, WI: Wisconsin Department of Public Instruction.*

Appendix 2: norms for TTR

APPENDIX 3
Coding Stuttering and Other Fluency Behaviors

A new set of codes has recently been developed for stuttering behaviors and will be included in all new issues of CLAN. The two new symbols are in Unicode and have shortcut combinations that work when you are transcribing in CLAN[13]. The complete list of stuttering/disfluency codes are:

Stuttering behavior	**Code**	**Example**	**Notes**
Prolongation	:	s:paghetti	Place after prolonged segment
Broken word	^	spa^ghetti	New code
Block	Unicode2260 ("not equal to" sign); *shortcut:* Hold F2 and =	≠butter	This example illustrates a block before word onset
Repeated segments	21AB (curly left arrow); *shortcut:* Hold F2 and /	↫r↫r↫rabbit OR like↫ike↫	Note that the curly left arrow brackets the repetition but leaves a recognizable target for mor
Whole word repetition	follow word with [/]	butter [/] butter	no < > necessary if only one word is repeated
Multiple whole word repetition	indicate number of repetitions in brackets	butter [x 7] x space N	Indicates that the word 'butter' was repeated seven times
Typical Disfluencies			
Phrase repetitions	<> [/]	<that is a> [/] that is a dog.	
Phrase revisions	<> [//]	<what did you> [//] how can you see it ?	
Filled pauses	-&	&-um &you_know	Note: multiword fillers should be connected with an underscore to avoid wrong word count

Tallying stuttering behaviors
Using the above codes insures that when you transcribe, and analyze your transcript, neither FREQ nor MOR (the syntactic program) will include filled pauses as words (stuttered words will be counted as intended words, which is correct).

SEE PAGE 29 for the current command to tally disfluencies.

[13] **In CLAN**, to see short cuts for any phonetic, prosodic or fluency codes (in case you lose this ☺), go to WINDOW → Special Characters, and a drop down menu will appear that you can scroll through to find the short cut. Please note that these key stroke combinations work inside CLAN and are **not** the key combos for typing in Word or other programs.

To get %stuttered words, FREQ your file, and then the **output of the fluency tally** can be put over the **token total** to give you %blocks, %prolongations, etc.

To tally fluency behaviors, CLAN uses a file that looks like this [fluency-sep.cut], in the LIB folder, inside its Fluency folder, with these symbols, one per line[14]:
&-*
↫
≠
^
⁚
[/*] or </*> depending on whether you are counting occurrences of repetitions or words themselves.
[x *]

The only reason why we are reprinting it here is that some people want to either track or NOT track certain behaviors. You can add a behavior, such as hesitations, or fillers, by simply adding lines to this file and saving it. You can also remove a behavior if you don't think it is relevant to your fluency work.

NOTE: Remember that you can insert the symbol codes from the Window drop down option list using "special characters"; if you are typing or inserting and the block or repetition code enters as something other than what you want (for example, a box, rather than the curly arrow), you need to set your default CLAN font so it accesses Unicode. Go to View, and set font, and click a specific font. The best option is Arial Unicode or CA font, but if you don't see those, any font followed by Unicode (Lucida) will do. If you don't see ANY Unicode font, go back to the CHILDES site, where you can download the font file and save it to CLAN lib.

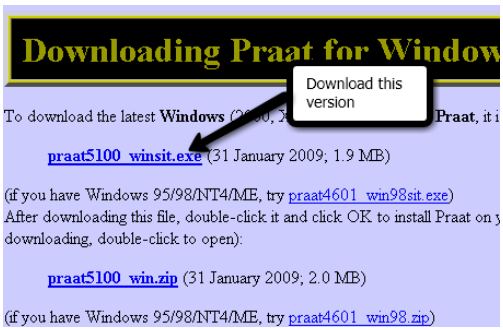
[14] The asterisk is very powerful, and is a wild card that allows the program to find the code regardless of what letter combination it is next to. CLAN uses a lot of asterisks, as do many computer programs, and search routines, such as those in PubMed, EBSCO, etc. Try it if you haven't ever done this.

ACOUSTIC ANALYSIS

One of the more amazing talents that CLAN has is the ability to go straight from a linked transcript (audio or video) to acoustical analysis that many clinicians would probably gladly do if it were easy enough (to look at properties of sound segments, intonation contour of utterances, rate calculations, etc. It is a completely seamless interface. This next section shows how to do this.

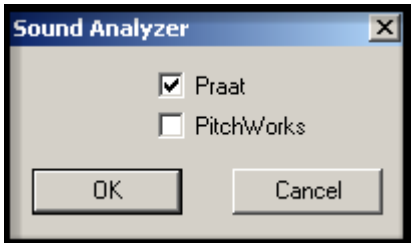
PRAAT

In order to run the acoustic analysis, you will need to download PRAAT. This can be done at http://www.fon.hum.uva.nl/praat/ . Click "Windows" (or MAC) in the upper right hand corner and then click on the first download option on the next page. Instructions for how to install are listed on that page. There is also a PRAAT tutorial for CLAN in the tutorial series at the CHILDES web site (file 13; PRAAT pdf); you should probably read that as well.

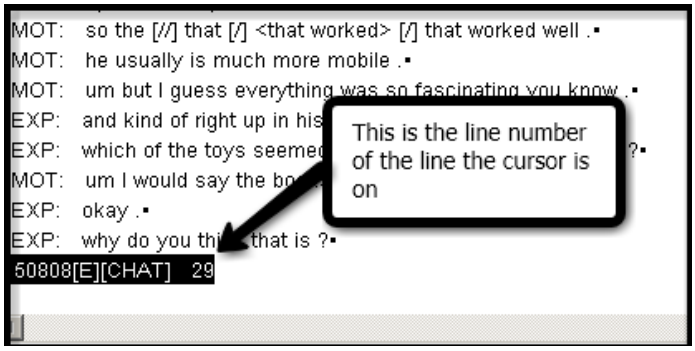


From CLAN to PRAAT

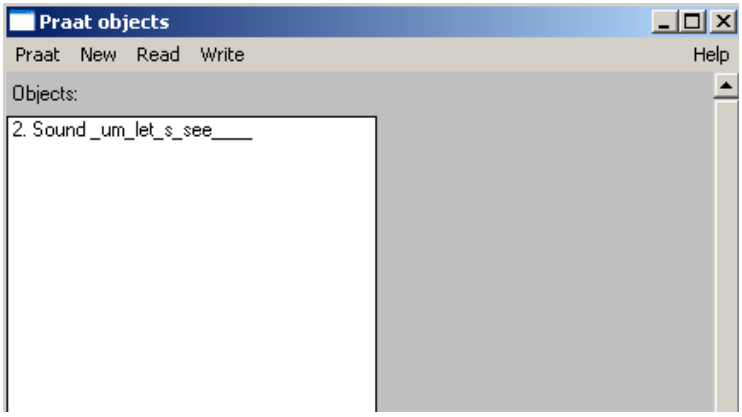
The bullets in CLAN can be sent directly over to PRAAT, but you'll have to set it up to do this. First, open both programs. PRAAT MUST be open first in order to use it with CLAN. When you open PRAAT, you can close the window labeled PRAAT Picture and leave open the window labeled PRAAT Objects. In CLAN, you can close the Commands window. Then, in CLAN, go to Edit → Select Sound Analyzer. In the new window, click the box next to PRAAT and make sure the other box is unchecked. Once you have set this, the setting will remain this unless you change it. You don't need to do this step every time.



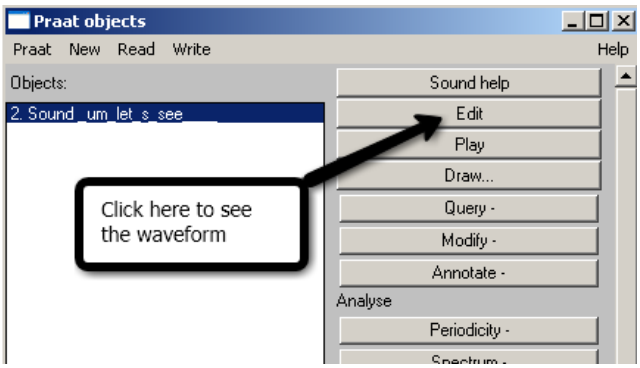
Next, open the transcript you are working on. Using the spreadsheet, look up each matched target. The spreadsheet will tell you in which transcript and on what line the target is located. The line numbers in CLAN show up at the bottom of the screen.



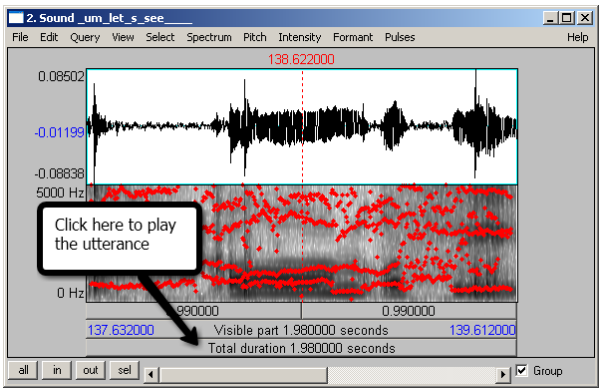
Locate the line containing the target and click your cursor on that line. Then, go to Mode → Send to Sound Analyzer. For this to work, PRAAT must already be open. Once you do this, the text from that line in the transcript will show up in the PRAAT Objects window.



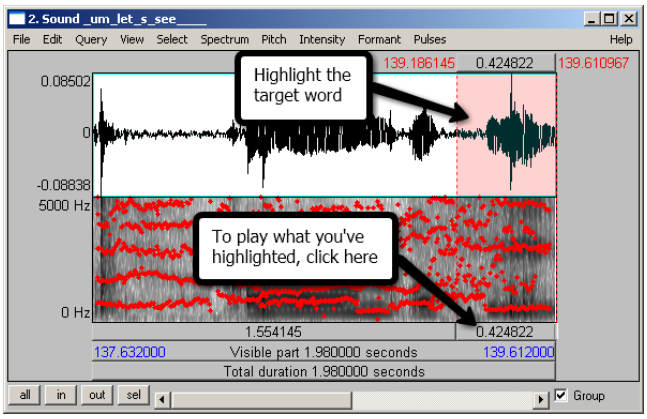
Click on the object you are analyzing and then options will show up on the right side of the window. Click Edit to see the waveform and spectrogram.



Another window will open and in that window you will see two version of the same thing. The top one is the waveform and the bottom one is the spectrogram. If you click on the bar on the bottom, the whole utterance should play.

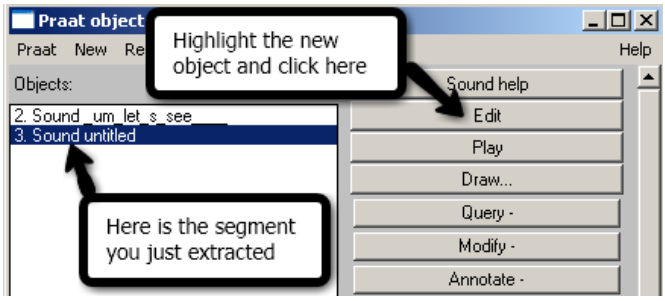


After listening to the utterance, find the section of the waveform that contains the target word and highlight it by clicking at the beginning of the word and dragging your mouse over to the end of the word. Play what you have selected by pressing on the top bar. Adjust your highlighted portion until you are satisfied that you have the whole word and nothing else. This will take some practice.

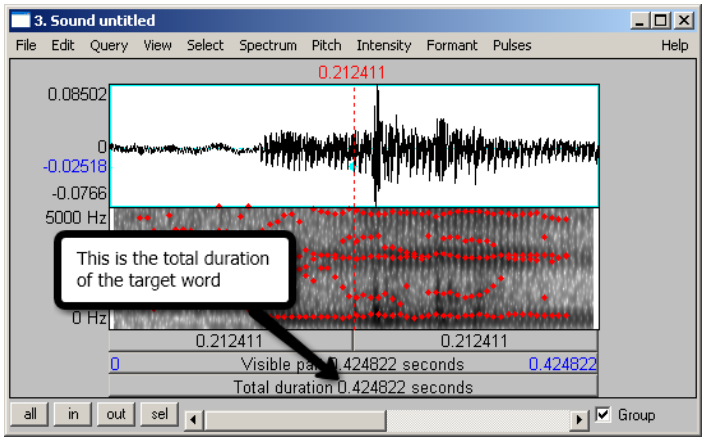


Word (or utterance) duration

Once you have isolated the target word, go to File → Extract Selected Sound (Time from 0) and close the window. This will create a new object in the Objects window. To look at a close-up version of the target word, click on the new object and click Edit.



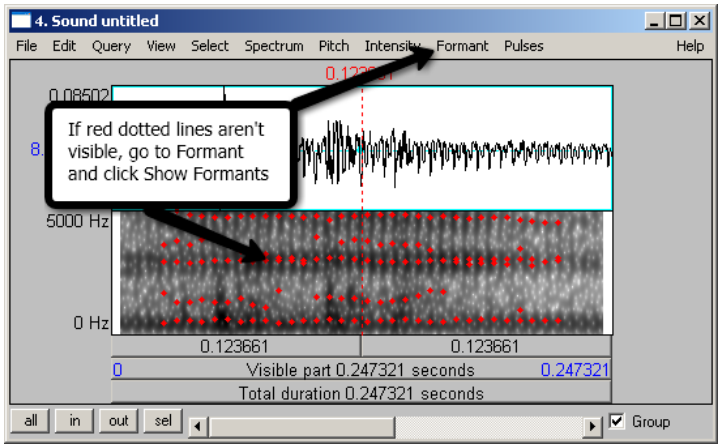
When the new waveform window opens, record the total duration of the word in seconds. You can find this value at the bottom of the window where it says Total Duration.



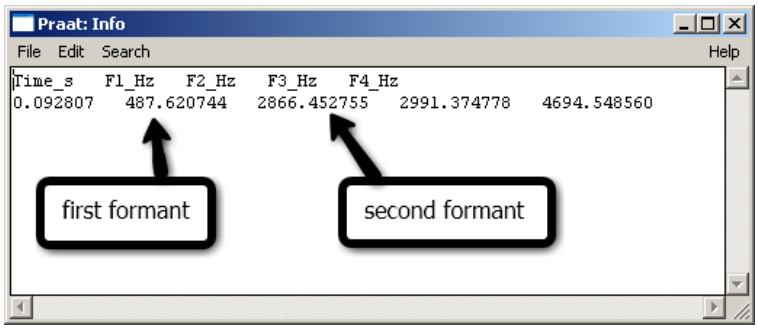
To do a much more precise measurement of speaking rate than you can with a stopwatch (☹), you can mark the beginning and end of an utterance, get its timing across the bottom window, count the number of words/syllables in the utterance (you already have them written out!), and there you go!

Formants

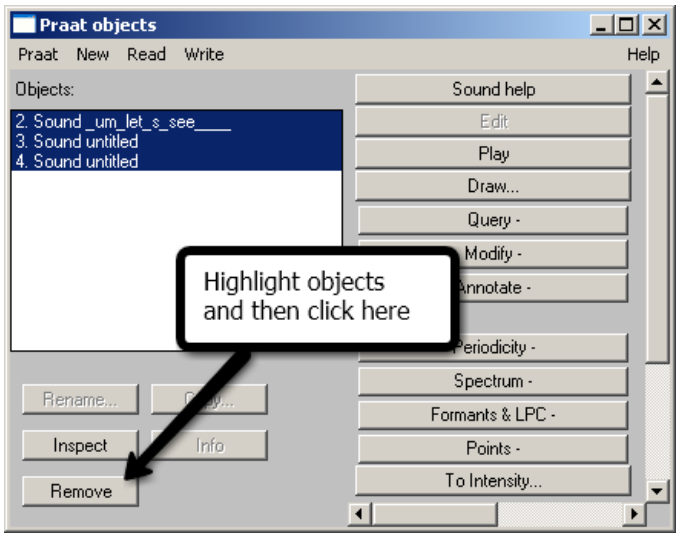
To find the first and second formants, look for red dotted lines in the spectrogram of the isolated vowel. If they are not there, go to Formant → Show Formants. This will make the dotted lines show up. Then, find a place where the formants look as stable as possible. This means they should make a straight line across a section of the vowel. Click in the middle of the most stable part.



Then go to Formant → Formant Listing. A new window will open. In this window, the formants are listed. Record the first and second formants in the appropriate columns in the spreadsheet.



Once the spreadsheet is filled in for that target, you can move on to the next one. In order to keep things uncluttered to avoid any confusion, you should remove the objects you are finished with from the Objects window before starting to work on a new utterance. To do this, highlight the unwanted objects and then click Remove at the bottom of the window.



Other typical PRAAT options are to show Intonation Contour, Intensity, etc. There are a number of PRAAT tutorials on the web for you to explore. It is possible to write scripts for some acoustic analyses that will automate the process. This is an advanced topic, but might be worth learning if you have lots of files that you wish to evaluate.