

ANNOTATED SYLLABUS

1

STUDY OF LANGUAGE LAB
SYLLABUS
WILLIAM & MARY

PEOPLE

PROFESSORS:

Kaitlyn Harrigan

Anya Hogoboom

Leslie Cochrane

TEACHING ASSISTANTS:

TA for each workshop section

POLICIES

ABSENCE POLICY: Attendance is expected at every lab meeting. If you are unable to attend, email your section's TA (not your professor) for instructions on the make-up work (which is often in addition to what students attending the workshop have to turn in). You must turn in the make-up work by the start time of the next week's workshop to receive credit. You may do this up to two times with no penalty. If you will need to be absent from the workshop more than twice, you will need to make special arrangements with the professor of your Study of Language section.

ACCOMMODATIONS: If you have a disability or if there is any circumstance that may impact your participation in the workshop, please let your professor know as soon as possible so we can discuss the best way to meet your needs in this class. Student Accessibility Services also provides accommodations for documentable disabilities.

EQUIPMENT: Necessary equipment for each week's lab is listed on the schedule; details may also be announced via Blackboard. Some workshops require you to bring headphones, for instance, or to download software ahead of time. When a laptop is listed on the schedule, you must bring a laptop that is fully charged (it is unlikely you will have access to an outlet) and that is an actual computer with a Windows or Mac operating system (*not* a tablet, such as a Chromebook or an iPad). If you don't have a laptop, you can check one out from the Media Center.

ASSESSMENT: You will receive full credit for labs in which you are *actively engaged for the full time period and upload or turn in any required work* (or for which you complete the appropriate make-up work on time). Lab activities are designed to be finished during class. Labs are 10% of your final grade in your Study of Language Section.

ANNOTATED SYLLABUS

Full Credit - Grading Percentage

12 labs - 100%

11 labs - 90%

10 labs - 80%

9 labs - 70%

8 labs - 60%

7 labs - 50%

6 labs - 40%

5 labs - 30%

4 labs - 20%

3 labs - 10%

1 or 2 workshops - 0%

SCHEDULE

Be sure to prepare and to bring all materials listed for each date. Changes to the schedule will be announced over Blackboard.

Week 0 **Hypothesis testing: expletive infixation**

Bring: writing utensil

Week 1 **Phonetics: discovering unconscious sound properties**

Prepare: download Excel and Praat following instructions on Blackboard

Bring: charged laptop (with Excel and Praat), headphones/earbuds, writing utensil

Week 2 **Sociolinguistic data collection: correlations between linguistic & social variables**

Prepare: collected data due on Wednesday before Friday workshop

Bring: charged laptop, writing utensil

Week 3 **Phonological processing: complex perceptual cues**

Bring: charged laptop and headphones/earbuds, writing utensil

Week 4 **Cross-linguistic investigation: phonetic markedness**

Bring: pencil with eraser; optional highlighters

Week 5 ***Language Revitalization: documentary and discussion***

Bring: writing utensil

ANNOTATED SYLLABUS

- Week 6 **Morphological processing: pseudo-affix effects**
Bring: charged laptop, writing utensil
- Week 7 *Break – no workshop meeting*
- Week 8 **Cross-linguistic investigation: syntactic typology**
Bring: charged laptop, writing utensil
- Week 9 **Syntactic processing: garden-path effects**
Bring: charged laptop, writing utensil
- Week 10 **Language acquisition: discovering productivity in child speech**
Bring: charged laptop, writing utensil
- Week 11 **Language variation: mapping American English dialects**
Bring: 6 different colors of pen, highlighter and/or colored pencil
- Week 12 **The linguistics of writing**
Bring: writing utensil
- Week 13 **Discourse analysis: narrative structure**
Bring: 6 different colors of pen, highlighter and/or colored pencil
- Week 14 *Last day of class – no lab meeting*

SAMPLE LAB #1

Language variation: mapping American English dialects

Some Background:

As linguists, we know that all languages have varieties, often known as dialects, which differ in their phonology, morphology, syntax, and lexicon. All language varieties are equally systematic, complex, and expressive. Varieties are associated with social factors, such as regional origin, race/ethnicity, and socioeconomic class. In this workshop, we will be identifying varieties of English in the United States that are associated with regions.

The data come from a survey of native English speakers in cities across North America. Participant's names, ages, and occupations are included in these data. In a survey over the phone, participants were asked questions that aimed to elicit particular linguistic features and demographic information, as well as asked to speak spontaneously for a stretch so linguists could record and transcribe their pronunciation. We'll be examining a small selection of the features recorded. This survey focused specifically on sound features.

Part 1

1. On the data sheet, examine Feature 1. On the map, use **color 1** to put a small **circle** around the dot for all the cities in which a participant's variety has the low back merger. Use a small **square** for partial mergers. Mark the cities of participants with separate low back vowels with an **X**. At the bottom of your map, write 'low back merger' and a circle in color 1 and 'partial' and a square.
 - The low back merger occurs when /ɔ/ and /ɑ/ have merged so that words like *cot* and *caught* are pronounced identically. This process happens over time: it is complete in some varieties but still in progress in others, which is why we are marking for partial mergers and for completely separate low back vowels.
2. Examine Feature 2. Using **color 2** put a small **circle** around the dot for all the cities in which a participant's variety shows monophthongization. At the bottom of your map, write 'monophthongization' and a circle in color 2.
 - Monophthongization occurs in words like *wide* when the vowel is pronounced as a monophthong rather than a diphthong: closer to [wad] than [waid]. The vowel also becomes slightly fronted. This happens in some areas preceding voiced consonants and in others preceding all consonants, as in *rice*. There are also borderline areas where it occurs only before /n/, /ɹ/, and /l/ as in *iron*.
3. Examine Feature 3. Using **color 3** put a small **circle** around the dot for all the cities in which a participant's variety shows fronting of /u/. At the bottom of your map, write 'fronting of /u/' and a circle in color 3.

SAMPLE LAB #1

- Fronting of /u/ in words like *boot* and *two* is common in North American English in all but a few regional varieties. In most varieties, fronting occurs most often after alveolar consonants. In our data, this feature is noted only when it occurs most strongly. In these cases, /u/ as in *dude* is almost as far front as /i/ as in *deed*.
4. Examine Feature 4. Using **color 4** put a small **circle** around the dot for all the cities in which a participant's variety has the northern cities shift. Use a small **square** for partial shifts. At the bottom of your map, write 'northern cities shift' and a circle in color 4 and 'partial' and a square.
- The northern cities vowel shift is a chain shift, in which several vowels are shifting over time. This process is complete in certain regions but still in progress in others. In a complete shift, for example, *hot* would be pronounced more like [hæt] than [hat].

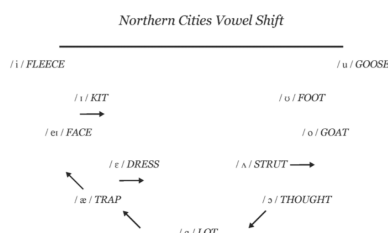


Image: graphic showing the progression of the northern cities vowel shift

5. Examine Feature 5. Using **color 5** put a small **circle** around the dot for all the cities in which a participant's variety shows fronting of /o/. At the bottom of your map, write 'fronting of /o/' and a circle in color 5.
- The fronting of /o/ occurs when the inherent diphthong [oo] is pronounced with the first vowel shifted forward toward the central space of the schwa (except when followed by a liquid). For example *coat* would be pronounced more like [kəʊt] than [koot].
6. Examine Feature 6. Using **color 6** put a small **circle** around all the cities in which a participant's variety is non-rhotic (rhotic is another word for [r]-like sounds). At the bottom of your map, write 'non-rhotic' and a circle in color 6.
- Non-rhotic dialects pronounce [r] after a vowel as part of the vowel; these occur in the British Isles as well as North America. A stereotypical example of this is the pronunciation of 'Harvard Yard' as [hɑ:vəd ja:d].
7. Draw a circle around the regions that have **all combinations of identical features** in common. Be careful – some cities have more than one feature marked (only encircle cities that have the exact same set of features) and some cities have partial features (which should be considered different from the feature itself). You may want to use contrasting colors where regions are close together. Note that it is possible for a regional variety to be an island within another region.

SAMPLE LAB #1

- The lines you are drawing between different areas are called *isoglosses*. Individual isoglosses can be drawn between individual features that vary. Here we are using a few *bundles* of isoglosses in order to draw a boundaries between regional varieties. However, these are always idealized boundaries, since regions are not always sharply defined (there may be transitional regions) and regional features can be layered (larger regions may be divided into smaller regions).

Part 2

1. On your map, label the regional varieties as follows:
 1. the variety that includes Denver: West
 2. the variety that includes Sioux Falls: North
 3. the variety that includes Fargo: North Central
 4. the variety that includes Chicago: Inland North
 5. the variety that includes Omaha: Midland
 6. the variety that includes Nashville: South
 7. the variety that includes Maryland: Mid-Atlantic (use arrows into the ocean)
 8. the variety of New York City: 'NYC' for New York City
 9. the variety that includes Albany: 'W.NE' for Western New England
 10. the variety that includes Boston: 'E.NE' for Eastern New England
2. For each variety, list below the (set of) feature(s) that characterize it:
 1. West:
 2. North Central:
 3. North:
 4. Inland North:
 5. Midland:
 6. South:
 7. Mid-Atlantic:
 8. New York City:
 9. Western New England:
 10. Eastern New England:

SAMPLE LAB #1

3. Which of the six features on your map do you notice in your own English speech:
 1. low back merger?
 2. monophthongization?
 3. fronting of /u/?
 4. Northern Cities Vowel Shift?
 5. fronting of /o/?
 6. non-rhotic?
4. What regional variety most closely matches yours? Is this a place you have spent a lot of time? If not, why do you think your speech resembles this variety?
5. On the map, why are there so many different varieties of English in the eastern US, compared to the western US?
6. Do regional varieties explain all English language variation in the US? For example, would we expect all the people in Williamsburg to speak English in the same way? If not, why not?

If you have time (not necessary to hand in):

7. Using your phone, tablet, or laptop, take the online New York Times Dialect Quiz (www.nytimes.com/interactive/2014/upshot/dialect-quiz-map.html). At the end of the quiz, there is a prediction of three places in the United States you might be from, based on features of regional English varieties. What were your three places?
8. Have you or someone you lived with growing up spent time there? If not, why might the quiz have predicted that you did?
9. Why does the New York Times Dialect Quiz ask about lexical and grammatical features as well as phonological features?

References: The data for this workshop were taken from the Atlas of North American English by William Labov, Sharon Ash, and Charles Boberg. The vowel chart is from Walt Wolfram and Natalie Schilling (2016). The NYT quiz uses data from Bert Vaux and Scott Golder (2002).

SAMPLE LAB #2

Morphological processing: pseudo-affix effects

Objective. Over the past several weeks, we have been talking about Morphology. This week, our goal is to think about how this system is actually used in the human being.

Some questions to think about:

- *How are words stored, accessed and recognized during language production and perception?*
- *What are the components of the process of building up and taking apart words during perception and production?*
- *How can experimental evidence inform us about morphology as a “rule-governed system”?*

In this lab, we will be setting up a short experiment and running it (on ourselves!). Each of you will set the experiment up on your computer run the experiment on yourself. You will extract your data and upload this to a shared google document. I will then download all the files and we will look at our data together as a class. You can follow along the data processing steps (included here) to look at your own individual data. We will use Microsoft Excel to look at our data.

Step 1: Set-up.

You will run the experiment on yourself using PsyToolKit, a web-based platform for running and analyzing psycholinguistic experiments. Then you will share your data in a class google document and we will look at the whole class' data together.

Go to the Psytoolkit website (<https://www.psytoolkit.org/cgi-bin/psy2.6.0/>); or google “Psytoolkit” and click on “web-based/login”. Create an account. You will need an email to do this (it can be your W&M email or another one, whatever you prefer). When you register, you will get a confirmation email with your password. You can change your password if you'd like once you login.

- On the Blackboard site, you will find a file called morph_lab_package.zip. Download that. Open the file (or right-click and then “extract all”) to unzip the file. You should now have a folder containing several files: a text file called ‘morphexp.txt’, a photo image called ‘instructions.png’, and another zip file called ‘ldt.zip’. **Do not unzip the ldt.zip file** (just don't click on it or do anything to it and it will remain compressed in the folder!). Make sure you know where this folder is—you will need to access the files in here shortly.

Step 2: Create a new experiment.

- Once you've created a login and received the confirmation email, login to your account.
- Click the link in the left menu to “Create new experiment”
- **We will use Method 2 to create our experiment.** Give your experiment a name, e.g., Morphology_Experiment
- On the PsyToolkit website, select “Choose File” to upload the uncompressed ldt.zip file. After the file is uploaded, click “Create experiment from an uploaded PsyToolkit experiment file”. That will take you a page where you can edit your experiment.

SAMPLE LAB #2

Step 3: Edit your experiment.

- In the folder of materials you downloaded, there is a text file called 'morphexp.txt'. Open this in a text editing program (if you have a Mac you should have one called TextEdit, if you have a PC you should have one called Notepad. Microsoft Word will not work for this). Copy all of the text, as is.
- Go to your experiment window on PsyToolkit. Delete the text in the window "Enter/edit in the box below:", and paste the text from morphexp.txt in that window.
- Click the yellow button that says "Save [code has been changed]"
- Next, we need to update the instructions for the experiment. In your folder, there is a file named 'instructions.png'. On PsyToolKit, click the button "view files" that is next to the "Save" button. Click the x button next to the current instructions.png to delete the default instructions. It will ask you if you really want to delete this file. Select "OK". Then, under the menu to "Upload files into this directory/folder" Upload the instructions.png that you downloaded from Bb. Choose that file and click "Upload". That should load the new instructions.png in the experiment files (You'll see a blue box with text for instructions). Now click the button "Go back to experiment".
- Finally, under the menu to "Compile and run", click "Compile". If you've followed these instructions carefully, that will yield the message "Successful compilation. You can now run the experiment."

Step 4: Run the experiment (read through all instructions in this section before you begin!).

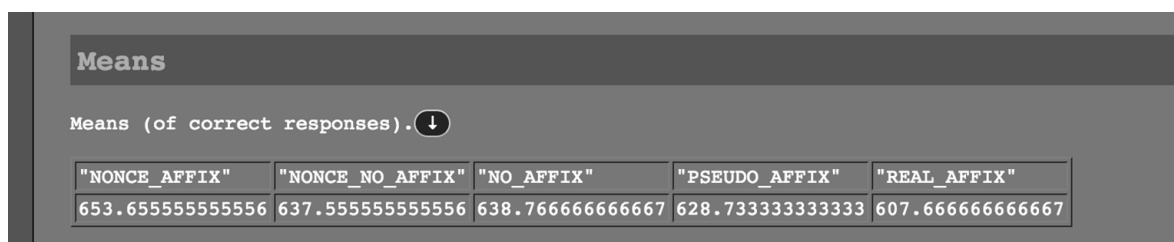
- You can now run the experiment on yourself!
- Click the button "Run experiment". That will take you to a screen with a red box that says "Click to start" As soon as you click that button, the experiment will begin, so make sure you are ready! The instructions will appear on the next screen, make sure you read them carefully!
- The experiment will take about 10-12 mins.
- At the end of the experiment, there will be a screen that says "Related words 0 ms Unrelated words: 0 ms Nonsense words: 0ms Press space to continue". **PRESS THE SPACE BAR** and then click the button that says **SAVE THESE DATA ON SERVER FOR LATER DOWNLOAD OR ANALYSIS**.
- That will take you back to the experiment window. If you've saved your data, you'll see a big table with the data.

Step 5: Extract your Data.

- After you've the experiment, you need to extract your data. Under the "Select a stored datafile" header, click the button that says "Select data file".
- Scroll down to the bottom (past your data!), where you see the header Set the meaning of the columns in the data file. And enter the following information, leaving the other fields blank:
 - Which column has reaction time: 6
 - Which column has status (correct=1)?: 8
 - Grouping variable (column)?: 4

SAMPLE LAB #2

- Click the button that says “Save parameters/run analysis on selected file”
- Next, scroll down under header “Analysis of multiple files” and click the button “Apply the analysis to all files stored here” (sometimes you have to click this twice!).
- You should see a table of means across conditions. It should look like this:



"NONCE_AFFIX"	"NONCE_NO_AFFIX"	"NO_AFFIX"	"PSEUDO_AFFIX"	"REAL_AFFIX"
653.655555555556	637.555555555556	638.766666666667	628.733333333333	607.666666666667

Image: screenshot of “Means” for reaction time across conditions

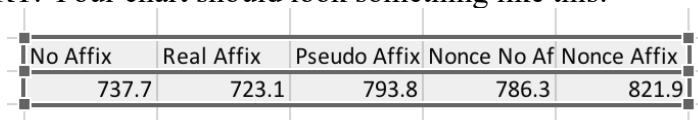
Step 6: Share your Data.

- Now share your mean response time for each condition so we can look at all the data.
- We will share our data using google documents. Open the google doc linked on Blackboard.
- Add your name and your mean RTs for each condition (make sure you are copying the right number—“Mean response time” and putting the right RT in the right column!)

Step 7: Analyzing the Data!

Now you will analyze your data. Once everyone’s data is uploaded, we will be looking at all the data together on the projector, and you can see whether your individual data fits the class profile! (Remember, the measures we use in psycholinguistics are often too nuanced to see much in the patterns of individuals... that’s why we run lots of subjects! This means you might not see much in your own data!)

We will be looking at our data in Microsoft Excel. We will basically be putting our data in a small chart and turning that into a graph. On the top line, type the name of each condition in each box. We have 5 conditions, so this should be across the first 5 columns. Below each condition, type (or copy in) your mean RT. Your chart should look something like this:



No Affix	Real Affix	Pseudo Affix	Nonce No Af	Nonce Affix
737.7	723.1	793.8	786.3	821.9

Image: screenshot of table from Excel showing reaction time averages across conditions

Notice that I have changed the order from the default order presented in Psytoolkit and our google spreadsheet. This is because I wanted them to show up in the order that it makes the most sense to talk about them. Make sure you are putting the right numbers with the right conditions! Once you have your chart, highlight the cells in the chart. Then go to the “Insert” tab. Here, you should have the option of different types of graphs. Click on the type of graph that you would like to represent your data! I like bar graphs, so that’s how I will be showing the class data. Now you should have a graph! From here, you can add a title and axis labels.

SAMPLE LAB #2

Morphology Workshop: Study Background.

Morphology is about words—how they are formed, stored and accessed. One model of the morphology system is that words are organized in the lexicon based on *roots*, and that morphemes are added productively to these roots based on morphological rules. This kind of model for organization of the lexical predicts that to access a word in the lexicon, we have to first identify the root. This predicts we go through a process of *affix stripping*—or identifying affixes and removing them to get to the root. So, this model predicts that when we get a word like *runner*, we have to first notice the morpheme *-er*, strip this off to discover the root *run*. Of course, we don't know until we've done this affix stripping process what the root is, or even whether the word is actually a morphologically complex word at all! To make things more complicated sometimes non-morphologically-complex words can initially *appear* to have morphemes, such as *father*. This theory predicts that when accessing a word like *father*, we would initially strip off the *-er*, and check for “fath” in the lexicon. At that point realize that we have no stored form “fath”, and that *father* is actually a root! Then we would need to reanalyze and search for “father”. Generally, in psycholinguistics, we understand more laborious processes to take longer. Reaction time is something we can measure, which means we can test our models of processing and lexical access by thinking about specific timing predictions that they make. In this case, we can test whether people actually seem to be doing an affix-stripping process by comparing reaction times for words like “runner” and words like “father”.

The Stimuli.

The stimuli for this *Lexical Access task* help us test the model for lexical access described above. In a lexical access task, participants see a string of letters on the screen and have to decide if it's a word or not. We assume that to do this, they must search their lexicon and “retrieve” the word if it's actually a word. If not, they will reject the string. Our stimuli consist of 5 conditions—three real word conditions and two nonce (i.e. fake) word conditions. The conditions are listed below:

1. **No Affix:** these are free morphemes that do not appear to have any affixes. Examples are things like *steak* or *mountain*.
2. **Real Affix:** these are morphologically complex words. The words we've used here contain only one affix. Examples are things like *runner* or *decode*.
3. **Pseudo Affix:** these are free morphemes that *appear to have morphemes* initially. Examples are things like *father* and *resume*.
4. **Nonce No Affix:** these are novel words that do not appear to have any affixes. Examples are things like *noogle* or *froum*.
5. **Nonce Affix:** these are novel words that appear to have an affix. Examples are things like *milger* or *desundle*.

SAMPLE LAB #2

A note about the methods: There are many factors to think about when you are creating stimuli like this! One thing we know about lexical access is that people are *faster* to access words that are more frequent. Since we are not interested in looking for frequency effects here, we want to make sure that we control for (i.e. match across conditions) frequency in our stimuli. This is to make sure that, for example, if we find that people are slower in the affix condition, it isn't just because the words we've chosen for that condition happen to be less frequent than the other conditions! We want to make sure we're really comparing words across the conditions fairly, changing only the relevant thing (in this experiment that is morphology!). I have controlled for frequency here by using a frequency corpus and matching the frequency of sets of words across our three real word conditions. So, each word in the pseudo-affix condition has a word that matches it in frequency in both the real-affix and the no-affix conditions. This ensures that the *average frequency* across each of our conditions is the same, and helps us be more sure that timing differences are really about the morphology and not something else! Can you think of anything else that I might have made sure to control for in this kind of experiment?

Questions for Discussion

1. What do we expect to see in terms of reaction times across each of the real and non-word conditions?

Real word conditions (rank in order of RT, slowest to fastest): no affix, pseudo affix, real affix

_____, _____, _____

Nonword conditions (rank in order of RT, slowest to fastest): affix, no affix

_____, _____

2. Why do we expect to see this? Explain in terms of the rule-based morphological system.
3. What did we find as a class? Were our predictions fulfilled?
4. How about your own results? Did they match the class' or differ?

SAMPLE LAB #3

Language acquisition: discovering productivity in child speech

In this Workshop, you'll be looking at a corpus of children's utterances and Child Directed Speech from a database of corpora called CHILDES.

Some Background:

In this workshop, we will be looking at some corpora from the CHILDES (**C**hild **L**anguage **D**ata **E**xchange **S**ystem) database, which is the largest database of child language and child directed language. It is an extremely useful tool for looking at patterns in language development and language input. CHILDES was established in 1984 by Brian MacWhinney and Catherine Snow. CHILDES includes data from many languages, and even bilingual families and some children with language impairments. There are several ways to browse the corpus—we will use the online browsable corpora, which we will access from the CHILDES website.

We will investigate one child's production of the morpheme *-ed*, which we use in English to mark past tense (i.e. *walk* → *walked*).

1. Getting into the corpus.

First we will browse some transcripts with audio/video included. To access the browsable transcripts, go to the CHILDES website: <https://childes.talkbank.org/> and click on the link for “Browsable Databases”. This will bring up a page with some information about using the browsable databases, as well as a list on the left-hand side of the different categories of databases that are available online. Click on the category called, “Eng-NA”, which will then display browsable transcripts from English speaking children in North America. Click on the Rollins database. You will see a list of the transcripts in this database displayed as links in the upper left-hand corner of your screen, all numbered and with the extension *.cha*. You will also see that some of them have video icons next to them, which look like this:



Image: screenshot of “play” button indicating video

Click on any of the transcripts with the video icon. To the right, you will see the transcript, and on the bottom left, you will see the video that is the source of the transcript. Press play, and watch a few minutes of the video, following along in the transcript. You can skip around if you want—wherever you click in the video, you will automatically jump to the corresponding spot in the transcript. Notice that there is a very specific format that gets used for this kind of transcription, which includes a lot of information about pauses, laughs, interruptions, etc. This can make it look a little cluttered at first, but after a few minutes you get used to reading through this! Most of this will not be important for our purposes today, but we do want to pay attention to one thing: to the left of each utterance, you will see a code representing the speaker. The most common ones are **CHI*, which stands for *child*, and indicates that the utterance came from the child subject, and **MOT*, which stands for *mother*, and indicates that the utterance came from

SAMPLE LAB #3

the subject's mother (who is often with the child during the time of recording). Other common codes are *FAT (*father*), *SIB (*sibling*) and *INT (*interviewer*). Sometimes codes for people's names are also used, like *SAH (*Sarah*). The Rollins corpus includes data from a longitudinal study, where the same children were recorded in parent-child dyad play sessions at 6, 9, 12, 18, 22, and 30-months of age.

Now go back to the list of corpora from North American English, and click on the Sawyer corpus. The data in this corpus were gathered from naturally occurring play encounters in a preschool classroom with approximately 24 3, 4, and 5 year olds, during their unstructured free play time. This corpus only has audio recordings and no videos. Click on one of the transcript links and listen for a few minutes, following along with the transcripts on the right side of the screen.

What are some of the challenges of transcribing naturalistic human language production? How do children make it especially challenging?

What do you notice about the difference between the linguistic productions of the children from the first corpus (Rollins) compared to the second one (Sawyer)?

2. Searching the database.

Of course the value of spending the time and energy recording and transcribing children's language input and production is that we can search these transcripts for different things and learn something about what children hear and produce. If we want to search the database for specific patterns, we can't do that through the browsable databases—we have to download them, along with a program created for searching the transcripts in this format. The program we use to analyze data in CHILDES is called CLAN (Computerized Language ANalysis).

The Brown Corpus

The Brown corpus was collected by Roger Brown between 1962 and 1966. Adam's family is African-American, middle class and well educated. He is a speaker of Standard American English. There are 55 files in the Adam corpus, collected while Adam was between the ages of 2;3 to 5;2. You can download the Brown corpus here: <https://childes.talkbank.org/access/Eng-NA/Brown.html>. Make sure to unzip the file. If you open it up, you will see a couple weird files and three folders: Adam, Eve, and Sarah. We are going to be working with Adam's transcripts, which are located in the Adam folder.

Installing and configuring CLAN for Mac OS X:

If you need to permit downloading of non-AppStore apps, go to SystemPreferences / Security / General, open the lock, and click on "Anywhere". Next, go to <http://dali.talkbank.org/clan/> and download the Mac version of CLAN. Click to open clan.dmg and then click to start the installer. It will install in your Applications folder.

SAMPLE LAB #3

Installing and configuring CLAN for Windows:

Go to <http://dali.talkbank.org/clan/> and download the Windows version of CLAN. (Current versions of CLAN are no longer compatible with Windows 95/98/ME.) CLAN will automatically install in c:/TalkBank and your working directory will be c:/TalkBank/CLAN/work. The installer will create shortcuts for CLAN and the /work folder.

You should now be able to click on the CLAN icon and open the program. You should see a small “Commands” window that looks like this:

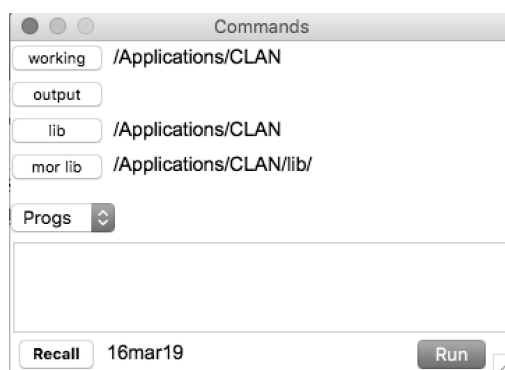


Image: screenshot of “Commands” window from CHILDES program

If the window does not open automatically, then type Control-d (Windows) or ⌘-d (Mac). This window controls many of the functions of CLAN. The main components of the Commands window are the command box in the center and the several buttons.

Setting your Working Directory:

The first thing you need to do when running CLAN is to set the working directory. The working directory is the place where the files you would like to work with are located. We will set the “Adam” folder inside of the “Brown” folder as our working directory. To set the working directory, press the “working” button in the Command window and use the navigation facility to locate the Adam folder. Press the Select Current Directory button. After selecting your working directory, you will return to the Commands window. The directory you selected will be listed to the right of the working button.

Using CLAN commands:

In this workshop, we’re going to be using a series of commands in the CLAN command window. The command has two parts: the first is the type of command, and the second is the file over which we are running that command. The command we are going to use is called “**freq**”, and it tells us the frequency of specific words in a transcript. The file we are running the command over will be in the form **.cha**. We will also restrict the searches we do in a couple other ways using buttons in the command window.

SAMPLE LAB #3

Let's practice with a simple case. First we're going to put in the command, then we're going to select the file.

- a) **Command:** click the “Progs” dropdown menu, and choose the `freq` option.

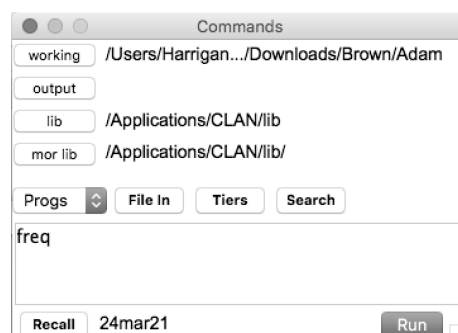


Image: screenshot of “Commands” window from CHILDES program with ‘freq’ in command line

- b) **File selection:** click the “File in” button. This will bring up a menu. In a left-hand box called “Please select files” you will see a list of all of Adam’s transcripts. Click on the top one (it should be called “020304.cha”), and move it over to the right-hand box called “Files for Analysis”. Now click “Done”.

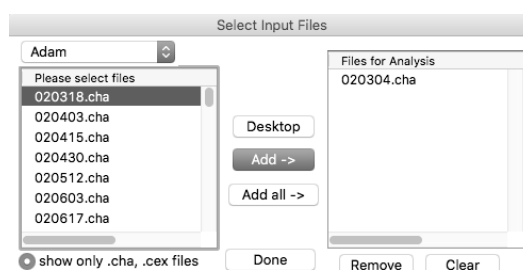


Image: screenshot of CHILDES window for selecting input files with one file selected for analysis

You should now see “**freq@**” in your commands window. Press “Run”. You should see a new window, called CLAN Output, open. The output gives you a list all of the unique words that were used by each speaker, along with the number of times that that word appeared in the transcript.

*How many times does Adam (remember that the target child is annotated as *CHI) use the word ‘tractor’ in the transcript that you looked at?*

SAMPLE LAB #3

3. Case study: past tense -ed production.

Now on to our case study! We're going to explore the production of the past tense morpheme, "ed" from the speech of one child—Adam. Our goal here is to learn something about how language is learned, as well as provide support for the kind of system we have been describing over the past semester. Consider two hypotheses about acquisition of the morphology system in children:

Hypothesis 1: *Children are learning language by imitation.* They are memorizing all words, including complex ones made up of multiple parts (such as happy-ness or un-believe-able), individually.

Hypothesis 2: *Children are learning rule-governed system.* They are hypothesizing a productive system for words, which includes word roots (happy, believe), affixes (-ness, un-, -able) and rules about how they combine.

What does Hypothesis 1 predict we should see (or not see!) in children's productions?

What about Hypothesis 2?

Because we are only interested in Adam's production, we are going to add in an extra constraint to our `freq` command in order to restrict our output to only Adam's productions.

Let's get our command set up:

- a) **Command:** choose the `freq` option from the dropdown menu.
- b) **File Selection:** click the "File in" button. This time, click "Add all", moving all of Adam's transcripts from the left- to the right-hand box. Now click "Done".

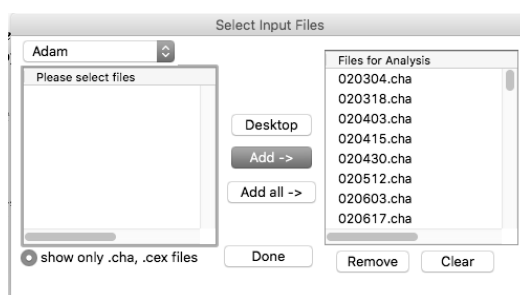


Image: screenshot of CHILDES window for selecting input files with all files selected for analysis

- c) **Restricting our search by speaker:** now click the "Tiers" button. When the "Tier Option" window pops up, click "More Choices" on the bottom. This will bring up a new window, that gives you options about restricting speaker Tiers. Set it so that only *CHI is selected, like this:

SAMPLE LAB #3

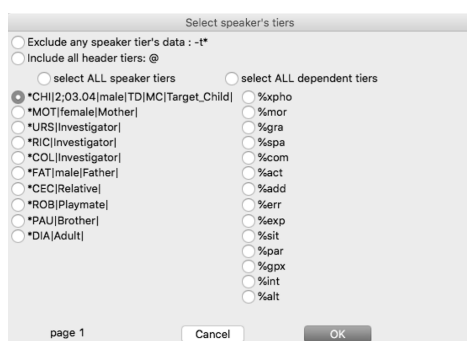


Image: screenshot of CHILDES window for selecting speaker “tiers”

If you do not see the “Tiers” button (this seems to be the case for the newer Mac version, CLANc), you can also just type or copy the next portion of the command in as follows: put a space after the @, then type (or copy): +t*CHI:

Your command should now look like this (whether you typed or used the Tiers button):
freq @ +t*CHI:

- d) **Restricting our search by morpheme:** now we’re going to type in the remainder of the command, to tell CLAN that we want it to look for the –ed morpheme. Type a space after CHI:, and then type (or copy) the following: +s"*ed "

(Make sure you include the space between the first part of your command and this addition, and make sure you have a space between the ed and the ")

Your Commands window should now look like this:

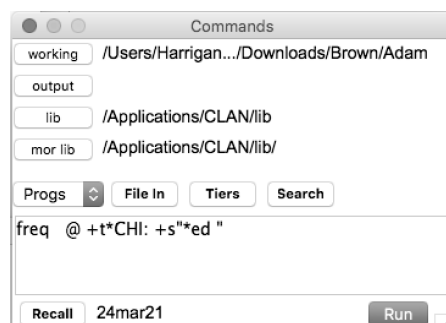


Image: screenshot of “Commands” window from CHILDES program with full script in command line

This command will find the frequency of each word that Adam uses across each transcript that ends in –ed. Now click “Run”. Your output window will now show you a list for each of Adam’s transcripts, including each word ending with –ed and the number of times it shows up in the transcript. It may take a minute or two for it to finish generating the output! (If it doesn’t work, double check all the steps above!)

SAMPLE LAB #3

get rid of them if you want) There are a lot of repeats, though, which we don't need. Highlight the column, click on the Data tab, and find the button that says "Remove Duplicates". Click this, and click "remove" on the menu that pops up. Now you should be left with a list of all of the unique –ed forms that Adam produced without any duplicates!



Image: screenshot of menu bar from Excel indicating the 'remove duplicates' button

How many unique –ed words do you have in your list?

Let's take a look at our list! The program isn't smart enough to know the difference between pseudomorphemes (*need!*), and real morphemes (*moved!*), so we will have to do a little sorting by hand. Read through your list and delete the words that are not past tense forms. Don't worry too much about the difference between true past tense inflections, past participles, and adjectives derived from verbs with a +ed rule. Although of course we *could* be strict about this if we had generated the context for where each of the words were produced, it doesn't really matter for our purposes today. If there is a verb version of the root of the word and it shows up in the transcript with an "ed" suffix, you can count it! Now you should have a finalized list! Read through your list and think about our hypotheses.

Which of our hypotheses are supported? Why? Which data points illustrate the process Adam is using?