Session 1719: Assessing fluency and language in children who may stutter: new diagnostic tools (Seminar 2-hours)

Nan Bernstein Ratner (University of Maryland; nratner@umd.edu)
Davida Fromm (Carnegie-Mellon University fromm@andrew.cmu.edu)
Brian MacWhinney (macw@cmu.edu)

TalkBank

The TalkBank System

DISCLOSURES:

- All three authors receive financial support for their work on the TalkBank project from the National Institutes of Health, NIDCD (**9R01HD082736-11; NIDCD: 1 R01 DC015494-01**) and the National Science Foundation (**NSF BCS-1626300/1626294).**

- Supplementary Disclosure: This session may focus on a specific approach, product or product line, tool, technique, service or model, and there may be limited or no information provided about other similar approaches, products, services, techniques, tools, or models.
- Description: We will describe TalkBank tools (KidEval, FluCalc) that have been funded by the NSF and NIDCD to be distributed free of charge via web interface, for both PC and Mac use. ***The authors do not receive any remuneration for SLP use of this software***.

FOR GUIDED SCREENCASTS OF HOW TO OBTAIN AND USE THIS FREE SOFTWARE, MAKE TRANSCRIPTS, AND USE KIDEVAL, GO TO http://talkbank.org/screencasts/

If you want the entire slide show,
just email us:

Nan Bernstein Ratner -- nratner@umd.edu
Brian MacWhinney -- macw@cmu.edu
Davida Fromm – fromm@andrew.cmu.edu

Why perform Language Sample Analysis (LSA)?

- Benefits of LSA (beyond the fact that it is required in some jurisdictions)
  - Ecological validity
  - Data not available from other sources
  - Easier goal-setting for intervention
- A speech/language sample is typically required in order to appraise children's fluency, as well.
  - We will discuss how to get both language and fluency measures from the same sample(s).

Problems with LSA

- Who has the time?
- Who has:
  - Memory for your old linguistics class?
    - Need to parse (mark morphemes) in the sample
  - Memory for the rules for complex routines such as **Developmental Sentence Score (DSS), Index of Productive Syntax (IPSYN)**?
  - Problems with math errors?
- Money for alternative programs (such as SALT)?

Problems with fluency assessment

- Accuracy of marking fluency in conversational samples is woefully unreliable (Ingham, Bothe and colleagues).
- This is partially a function of short-term memory during transcription, since the sample needs repetitive listening in order to mark the fluency after the sample is transcribed.
- The alternative is to do simple counting of disfluent events during the live sample (Lidcombe program procedures)
  - This is not ideal; it is subject to the same reliability concerns, and does not leave a unified record for review of the counts.
- Fluency specialists have yet to agree on whether to compute fluency over **words** or **syllables,** despite decades of discussion (Brundage & Bernstein Ratner, 1988; Yaruss, 2000)

Let’s compare some options:

1. SALT

http://www.saltsoftware.com/products/software

2. SUGAR

https://www.ncbi.nlm.nih.gov/pubmed/28738412

3. CLAN/KidEval

http://talkbank.org/manuals/Clin-CLAN.pdf

Comparing options:

	By hand	Systematic Analysis of Lang Transcripts (SALT)	Sampling utterances grammatical analysis – Rev: SUGAR	CLAN/Kideval
COST	?	$179	Free (uses Word)	Free
Coding	By hand	By hand	By hand	Automatic grammatical parser, only accuracy, disfluencies coded by hand
MLU w/m	?	BOTH	** MLU redefined & words per sentence	BOTH
Lexical Diversity	?	Total words, TTR, NDW, MTTR		Total words, TTR, NDW, MATTR, VocD
Mazes, fluency		Yes	No	Yes
Brown's morphemes		Most	Only –ing, -s (plural, poss, 3rd pers combined), -ed (past tense, adj combined; other (new) lexical affixes	All

Comparing options (continued)

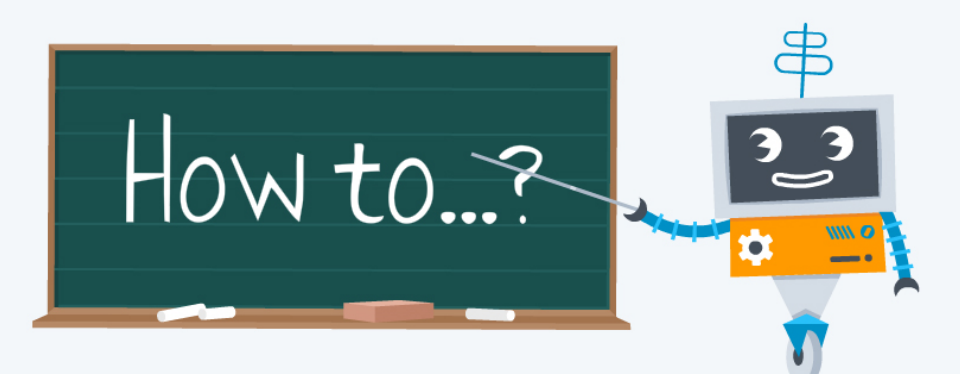
	By hand	Systematic Analysis of Lang Transcripts (SALT)	Sampling utterances grammatical analysis – Rev: SUGAR	CLAN/Kideval
Gramm Classes		**Yes**	**No**	**Yes via Eval**
IPSYN	laborious	No	No	Yes
DSS	laborious	No	No	Yes
Clausal Density	?	Yes	"Clauses per sentence"	Yes
Timing reports		Yes	Elapsed time to 50 utterances, no interlocuter	Yes, for each participant
Languages	Yours	English, Spanish	English	English, 10 other languages
Use Transcript for other purposes	No	No	No	Fluency analysis (FluCalc), EVAL, export to Phon, PRAAT
Custom queries		Limited	No	Yes
MEDIA LINKAGE		No	No	Audio or Video
PRE-DETERMINED FLUENCY CODES	No	Only mazes (typical disfluencies)	No	**YES**
FLUENCY COMPUTATION		Only mazes	No	**YES**

Comparison datasets (how to benchmark the child's language performance)

- Comparison database values:
- **SALT** -- free play (69 children to age 5;8);
- **SUGAR** -- conversation (385 children 3-7;11);
- **Talkbank/Kideval --** adult-child play (> 2070 children ages 1;8 -6;11).
  - Growing fluency database via NSF/NIDCD funding to FluencyBank
- Note: Output such as MLU, NDW, VocD, DSS & IPSYN have published reference values as well, reported in ***Clinician's Guide to CLAN*** appendices (free download).

Tell me more: How do you do this?

- Main steps (Overview… we will break this down):
  - Get the (free) program for Mac or PC: https://talkbank.org/
  - Instructions (free) – written manual: http://talkbank.org/manuals/Clin-CLAN.pdf ,
  - Instructions (free) -- tutorial screencasts in small steps: http://talkbank.org/screencasts/
  - Transcribe your sample from audio or video; or import typed materials or other formats by using CLAN TEXTIN command
  - Parse, insert Morphology using MOR
  - **Run KidEval (Language)**
  - **Run FluCalc (Fluency)**
  - **WRITE YOUR REPORT**



A LIVE Demonstration

to DEMYSTIFY the CHAT transcription process ...

Some basics to review:

- **Transcription basics**, in the CLAN editor window:
  - @Begin
  - @Languages: eng
  - @Participants: CHI YourCode, SLP Clinician
  - @Media: nameofmediafile, audio (or video)
- **Insert other case information: Tiers** ID Headers: fill in age, gender, other information as desired.
- **Transcript conventions/rules: Not very many!**
  - Basically, just type what you hear, no special coding necessary,
  - Avoid Caps except for I, proper nouns,
  - Do not use punctuation inside utterances,
  - Be sure you end utterance turns with punctuation (period, question mark, exclamation point, trailing off [+…].
  - Type intended target (rabbit) NOT pronunciation (e.g., wabbit).
- **Type @End at end of file.**

More basics:

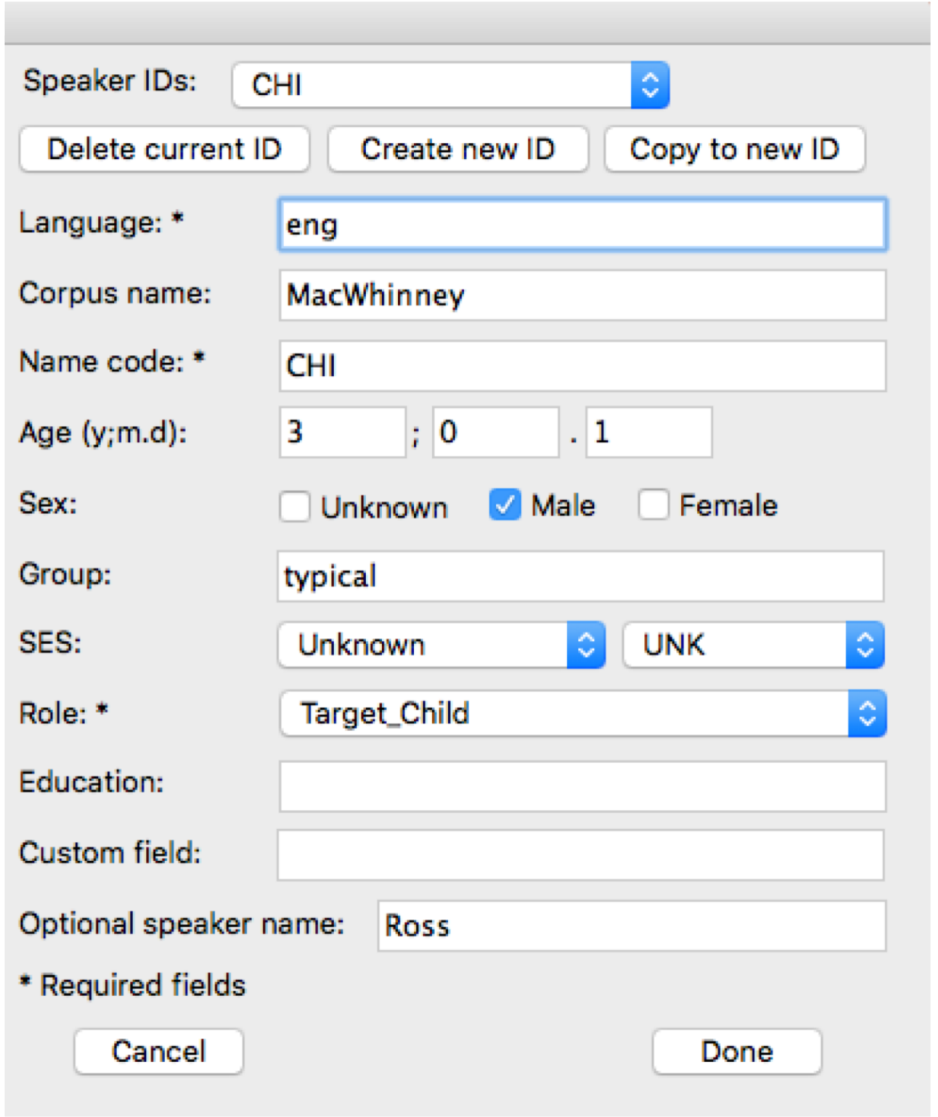
- **CHECK** your file
- **Get the English mor grammar** (under File option)
- In the **command window, type mor** and select the file you want to analyze
- When file has been morphemized, you can look up and fix any unrecognized words by typing mor +xb and using the same file – it will generate a list and location of words that were not found (usually transcription error). If not too many, you can skip this step and **go straight to KidEval or Eval or FluCalc.**

Adding clinical information to files

- We use the **header tiers** to add information to files:
  - AGE
  - GENDER
    - Both of these can trigger database comparisons to similar children
  - SES or maternal education
  - Time (such as pre- or post-therapy, baseline, etc.)
  - Group (such as Diagnosis, Therapy group [such as Social Language, period 5])

CLAN menu: Tiers → ID Headers

Input info (age, gender, etc.) in this drop-down field ...
and it automatically appears in the transcript header tiers

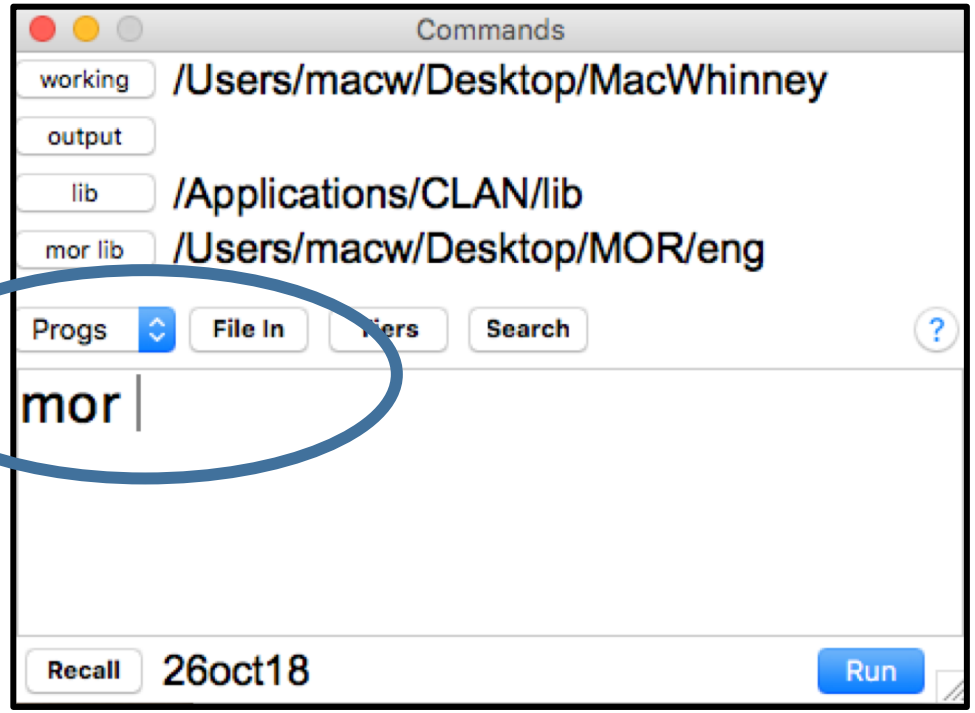


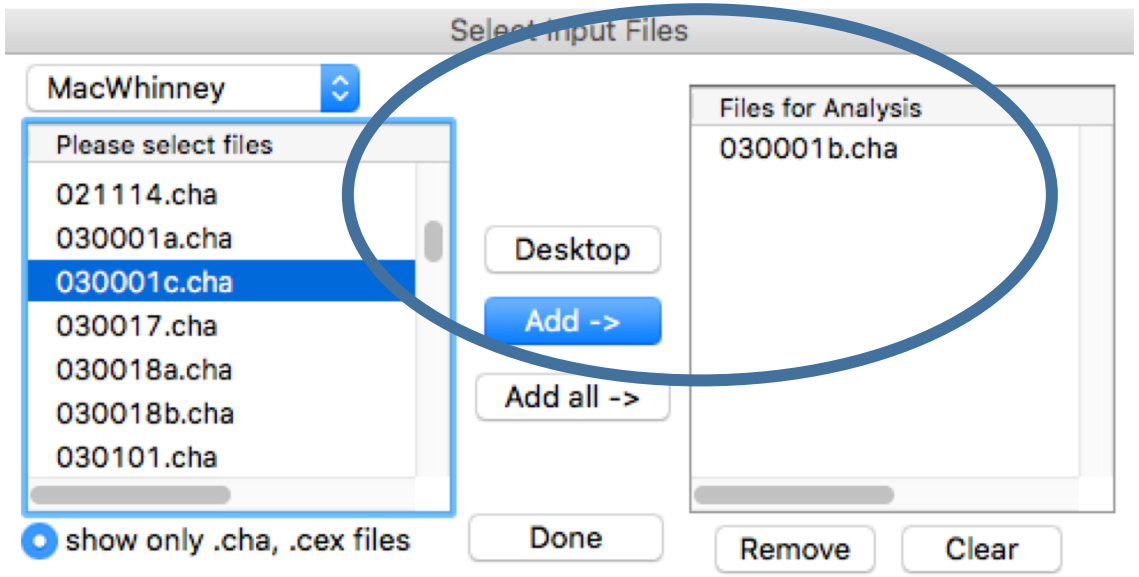

```
@Begin
@Languages:     eng
@Participants:  CHI Ross Target_Child , FAT Brian Father
@ID:    eng|MacWhinney|CHI|3;00.01|male|typical||Target_Child|||
@ID:    eng|MacWhinney|FAT||male|||Father|||
@Media: 030001b, audio
*CHI:   hot chocolate .
*CHI:   I just want one chocolate .
*FAT:   I'm just making it in here .
*CHI:   okay (.) and don't turn the heat on .
```

Run MOR command on file

1. Type **mor**, space, and press **File In** button

2. Select file from list, press **Add** button, and **Done**





VOILA!

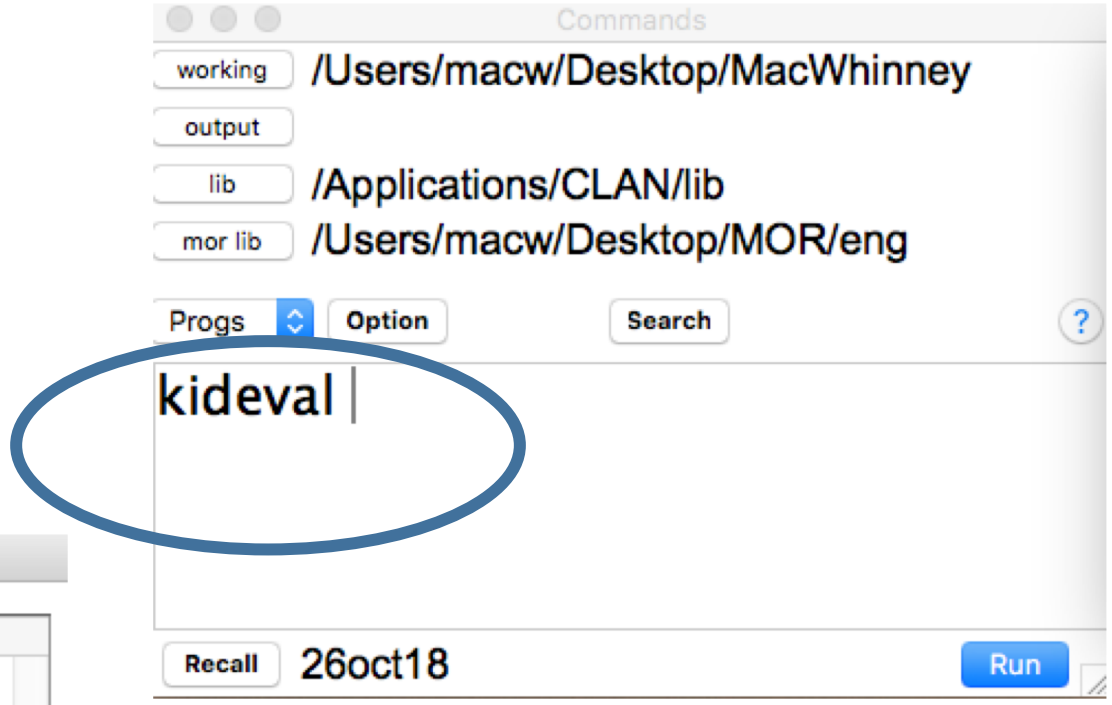
The transcript has parts of speech, morphosyntax, and grammatical relations in less than a second!

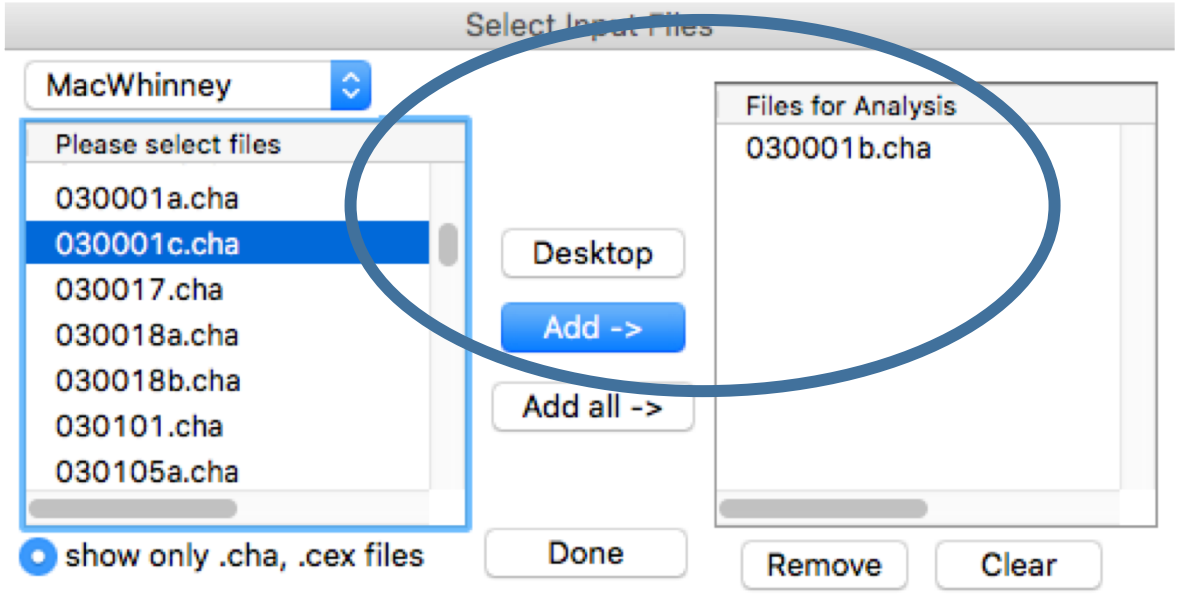
```
@Begin
@Languages:     eng
@Participants:  CHI Ross Target_Child , FAT Brian Father
@ID:    eng|MacWhinney|CHI|3;00.01|male|typical||Target_Child|||
@ID:    eng|MacWhinney|FAT||male|||Father|||
@Media: 030001b, audio
*CHI:   hot chocolate .
%mor:   adj|hot n|chocolate .
%gra:   1|2|MOD 2|0|INCROOT 3|2|PUNCT
*CHI:   I just want one chocolate .
%mor:   pro:sub|I adv|just v|want det:num|one n|chocolate .
%gra:   1|3|SUBJ 2|3|JCT 3|0|ROOT 4|5|QUANT 5|3|OBJ 6|3|PUNCT
*FAT:   I'm just making it in here .
%mor:   pro:sub|I~aux|be&1S adv|just part|make-PRESP pro:per|it prep|in adv|here .
%gra:   1|4|SUBJ 2|4|AUX 3|4|JCT 4|0|ROOT 5|4|OBJ 6|4|JCT 7|6|POBJ 8|4|PUNCT
*CHI:   okay (.) and don't turn the heat on .
%mor:   co|okay coord|and mod|do~neg|not v|turn det:art|the n|heat adv|on .
%gra:   1|5|COM 2|5|LINK 3|5|AUX 4|3|NEG 5|0|ROOT 6|7|DET 7|5|OBJ 8|5|JCT 9|5|PUNCT
```

Run KidEval command on file

1. Type **kideval**, space, and press Option button

2. Select file from list, press Add button, and Done





Your KidEval printout (Excel) -- cut into 3 parts

columns A-O

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	File/DB	Language	Corpus	Code	Age(Month)	Sex	Group	SES	Role	Education	Custom_field	Total Utts	MLU Utts	MLU Words	MLU Morphemes
2	MacWhinney/030001b.cha	eng	MacWhinney	CHI	36	male	typical	.	Target_Child	.	.	81	81	6.556	7.259
3	+/-SD	.	.	.	.	.	.	.	.	.	.	-0.901	-0.873	2.731	2.64
4														**	**
5	Mean Database	.	.	.	.	.	.	.	.	.	.	346.182	323.869	3.84	4.275
6	SD Database	.	.	.	.	.	.	.	.	.	.	294.326	278.086	0.995	1.13
7	Number files	.	.	.	.	.	.	.	.	.	.	137	137	137	137
8															
9	CURRENT COMPARISONS ARE PRELIMINARY AND SHOULD NOT BE USED FOR CLINICAL DIAGNOSIS.														
10	PLEASE SEE SLP MANUAL FOR RECOMMENDED INTERPRETATION.														
11															
12	+/- SD * = 1 SD, ** = 2 SD														
13	Database keywords: 3;-3;5|male														
14	Database date: 2018-04-27, 13:54														

Your KidEval printout (Excel) -- cut into 3 parts

columns P-Z and AA-AU

	A	P	Q	R	S	T	U	V	W	X	Y	Z
1	File/DB	MLU100 Utts	MLU100 Words	MLU100 Morphemes	FREQ types	FREQ tokens	FREQ TTR	NDW/100	VOCD D_optimum_average	Verbs/Utt	Word Errors	Utt Errors
2	MacWhinney/030001b.cha	81	N/A	N/A	158	552	0.286	50	38.93	1.198	0	0
3	+/-SD	N/A	N/A	N/A	-0.769	-0.725	0.852	0.075	-0.93	2.672	-0.575	N/A
4										**		
5	Mean Database	100	3.883	4.338	233.803	1329.365	0.222	49.445	50.168	0.679	0.92	0
6	SD Database	N/A	1.099	1.257	98.59	1072.584	0.075	7.36	12.077	0.194	1.6	0
7	Number files	120	120	120	137	137	137	137	137	137	137	137

	A	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU
1	File/DB	retracing[//]	repetition[/]	DSS Utts	DSS	IPSyn Utts	IPSyn Total	mor Words	*-PRESP	in	on	*-PL	*&PAST	*-POSS	cop|*	det:art	*-PAST	*-3S	*&3S	aux|*	~cop|*	~aux|*
2	MacWhinney/030001b.cha	0	0	49	N/A	72	N/A	552	15	4	4	11	14	0	26	46	4	0	4	3	4	10
3	+/-SD	-0.315	-0.59	N/A	0.973	N/A	0.591	-0.725														
4																						
5	Mean Database	14.343	14.088	50	7.989	100	86.622	1329.365														
6	SD Database	45.515	23.885	N/A	2.109	N/A	12.493	1072.584														
7	Number files	137	137	119	119	111	111	137														

(Partial) **List of measures in KidEval** (there are 44!)

- Total **number of utterances**
- **Mean Length of Utterance (MLU) in words** (for older kids, other languages
- **MLU in morphemes** (the traditional measure)
- **Types**: the number of different words used in the sample (from root form)
- **Tokens**: the total number of all words in the sample
- **TTR**: the Type-Token ratio, a measure of lexical diversity.
- Number of different words in 100 words (**NDW**): this value has evolving norms.
- **VocD**: another vocabulary diversity measure
- **Clause/utt: clausal density**, the average number of clauses per utterance

KidEval measures (continued)
All compared to normative sample

- **DSS: Developmental Sentence Score** (Lee, 1974).
- **IPSyn: Index of Productive Syntax** (Scarborough, 1990).
- Raw counts of **Brown's 14 morphemes** (in developmental order)

A	E	F	H	L	M	N	O	Q	R	S	T	U
Name	Age(Mo)	Sex	SES	Total Utts	MLU Utts	MLU- W	MLU-M	MLU100-W	MLU100-M	FREQ types	FREQ tokens	FREQ TTR
Kid1	32	M	2	374	321	2.321	2.548	2.38	2.71	182	919	0.198
+/-SD	.	.	.	0.216	0.103	-0.967	-1.019	-0.875	-0.823	-0.254	-0.177	-0.464

V	W	X	Y	Z	AA	AB	AC	AD	AE	AF
NDW/100	VOCD D	Verbs/Utt	Word Errors	Utt Errors	retracing[//]	repetition[/]	DSS Utts	DSS	IPSyn Utts	IPSyn Total
52	63.14	0.324	9	0	1	23	50	4.6	100	50
0.587	-0.032	-1.243	1.953	N/A	-0.227	0.522	N/A	-1		-1.913

AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU
mor Words	*-PRESP	in	on	*-PL	*&PAST	~poss|*	cop|*	det:art	*-PAST	*-3S	*&3S	aux|*	~cop|*	~aux|*
919	18	#	3	14	6	1	14	0	0	5	35	16	10	4
-0.177	-0.18	-0	-1	-0	-0.573	-0.451	-0.25	-0.896	-0.782	-0.3	-0.2	0.702	-0.398	-0.53

FIGURE 1. EXAMPLE OF CURRENT UT [truncated and screen wrapped]: please note that for many analyses, raw counts are also provided. In the event that a procedure (e.g., IPSYN, DSS) requires more eligible utterances than the sample provides, the child's data will not be used in the analysis of that LSA feature; N/A will be displayed.

Findings are referenced against large database and flagged if beyond 1 sd from mean

- TalkBank is currently funded through 2023 to provide norm-referenced values for children's language based on > 6,000 records from children
  - Conversational speech
  - Narrative
- CURRENT database contains thousands of children 0-6 years of age
  - Values in 6 month intervals
  - Can be separated by gender

Fluency Coding:

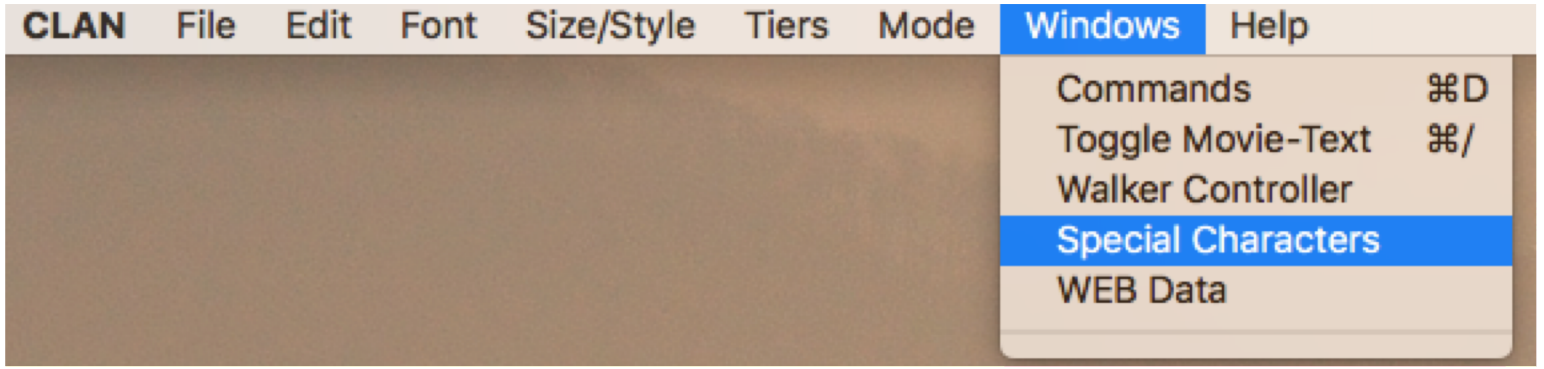
- New codes (will work for all languages) for stutter-like disfluencies
- Inserted through drop-down menu (no hunting for Unicode fonts)

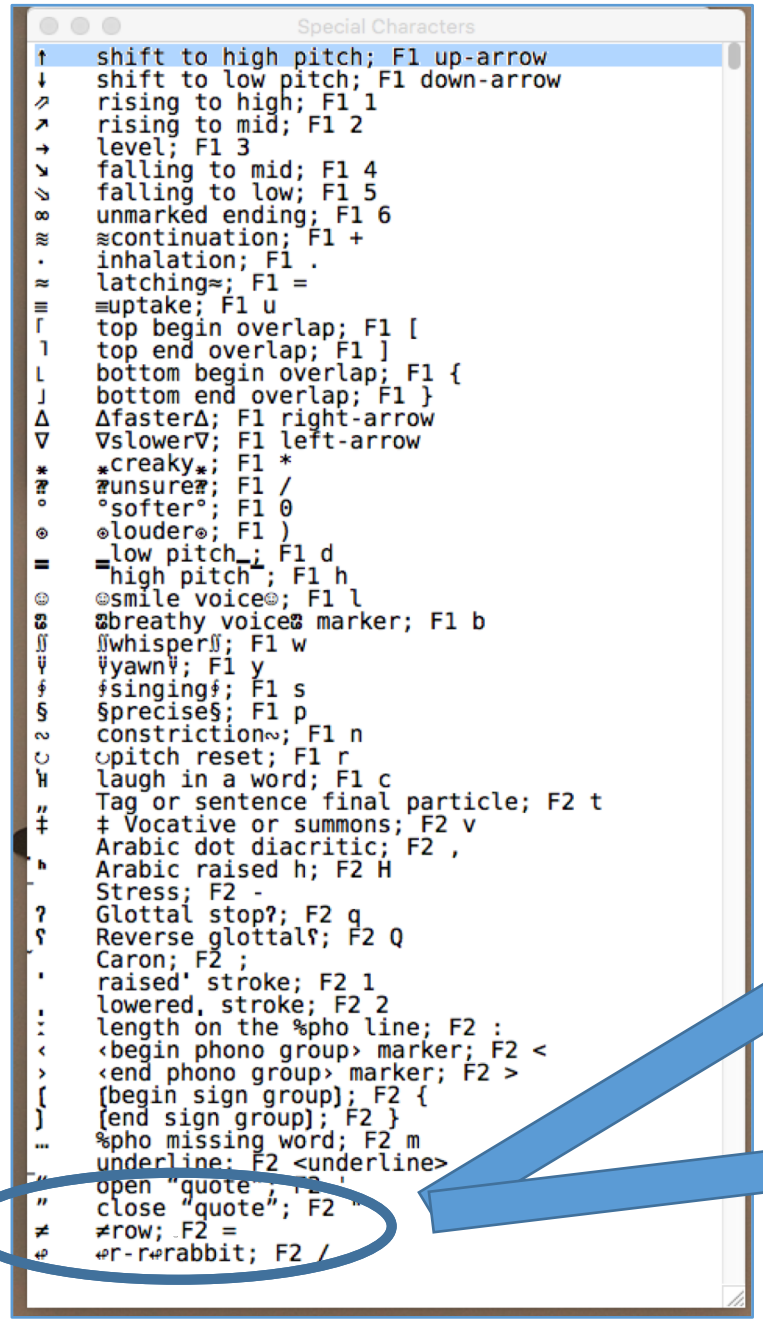
Stuttering behavior	Code	Example	Notes
Prolongation	:	s:paghetti	Place after prolonged segment
Broken word	^	spa^ghetti	New code
Block	Unicode2260 ("not equal to" sign); *shortcut:* hold F2 and =	≠butter	This example illustrates a block before word onset
Repeated segments	21AB (curly left arrow); *shortcut:* hold F2 and /	↫r-r-r↫rabbit OR like↫ike↫	The curly left arrow brackets the repetition but leaves a recognizable target for mor; iterations inside of the sequence are marked with hyphens
phonological fragment	&+	&+sn dog	Changes from "snake" to "dog"
other non-word strings	&	&gara	Word play etc.
Typical Disfluencies			
Whole word repetition	follow word with [/]	butter [/] butter	Repeated word counts once
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
Phrase repetitions	<> [/]	<that is a> [/] that is a dog.	Repeated phrase counts once
Phrase revisions	<> [//]	<what did you> [//] how can you see it ?	Revised phrase counts once
pause	(.) or (..) or (…)	(.)	Counts the number of short, medium, long pauses
pause duration	(2.4)	(2.4)	Adds up the time values, if marked
Filled pauses	&-	&-um &-you_know	Note: multiword fillers should be connected with an underscore to avoid wrong word count

Where IS that drop-down menu with the fluency codes?

In CLAN menu

- go to **Windows**
- select **Special Characters**





1. Find fluency codes at the bottom of the long list of **Special Characters**.

2. Double click on the fluency code you need and it will insert where your cursor is in the transcript.

or

Use the key combinations associated with the fluency code.

block on the word “row”

≠row; F2 =

repeated /r/ segment on “rabbit”

↫r-r↫rabbit; F2 /

Sample from a fluency transcript and video:

3 year 8 month old male, interacting with father

Look and listen for:

- repeated words do [/] do [/] do
- repeated segments ↫b↫because, ↫k↫ketchup, ↫th↫this, ↫v↫very
- sound fragments &+y &+y &+y
- prolongations h:ard
- unintelligible segment xxx

FluCalc

- Can be run in WORDS or SYLLABLES
- Tallies typical and stutter-like disfluencies (SLDs) SEPARATELY
- For children, computes Yairi & Ambrose's formula to numerically distinguish between typical and stuttered speech.

Run FluCalc command on file …

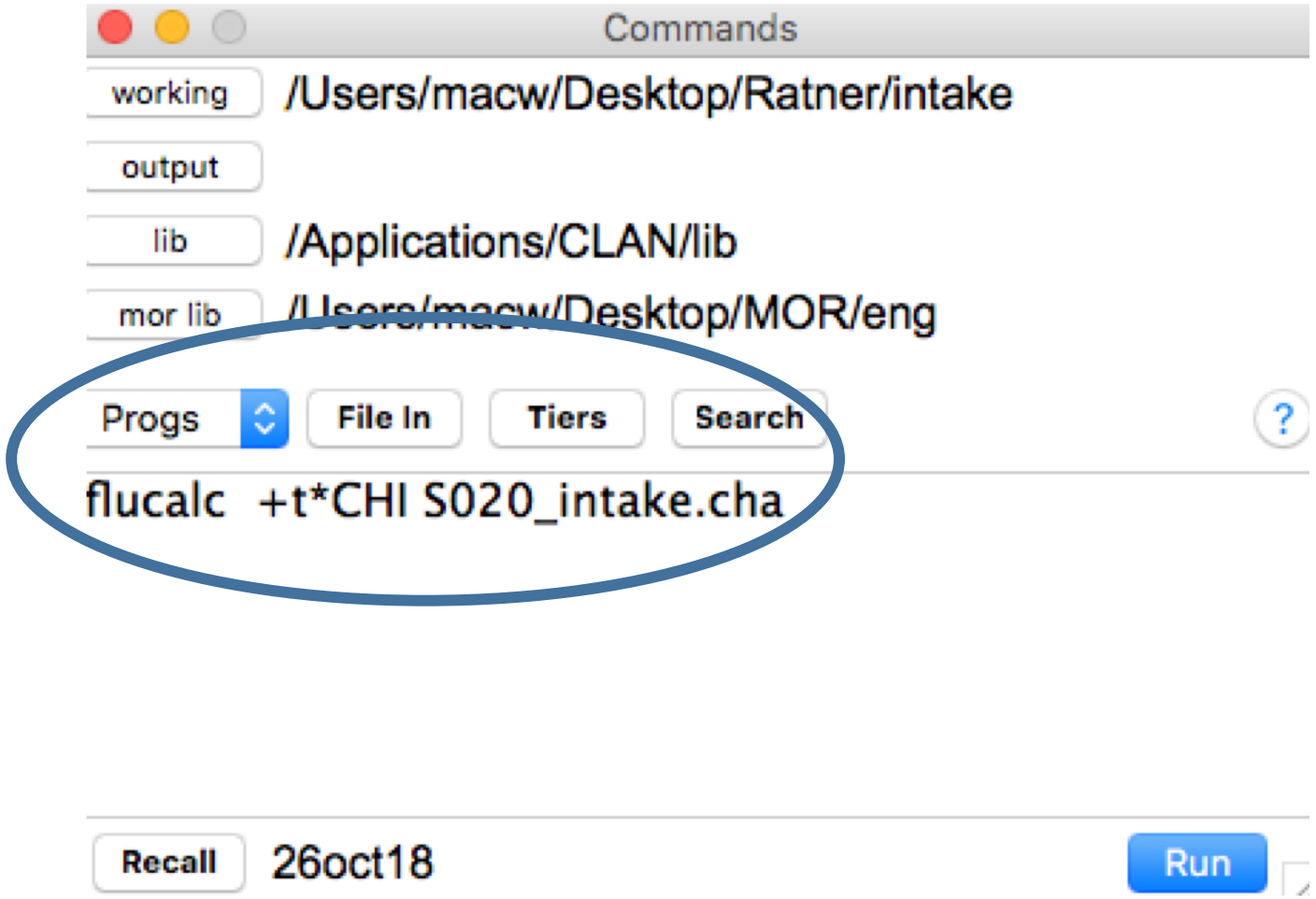
same process as previous commands

1. Type **flucalc**, space

2. Press Tiers button to select speaker OR type +t*CHI

3. press **File In** button, select file from list, press **Add** button and **Done** OR type in filename

4. Press **Run**



Sample FluCalc output

A	E	F	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
File	Age(M	Sex	mor U	mor Wo	# Prolongati	% Prolongati	# Broken wo	% Broken wo	# Blo	% Blo	# PW	% PWR	# PWR R	# WW	% WW	# WWR	Mean R
S001_i	2;8.	male	112	339	2	0.006	1	0.003	1	0.003	4	0.012	9	0	0	0	2.25
S002_9	3;8.	male	351	1432	14	0.01	1	0.001	3	0.002	81	0.057	107	47	0.033	52	1.242
S003_i	2;11.	male	71	199	0	0	0	0	1	0.005	5	0.025	7	0	0	0	1.4
S005_3	2;7.	male	316	842	5	0.006	1	0.001	7	0.008	23	0.027	30	22	0.026	28	1.289
S006_6	3;7.	male	91	237	0	0	0	0	0	0	0	0	0	1	0.004	3	3
S007_14	4.	male	127	346	3	0.009	1	0.003	2	0.006	21	0.061	33	7	0.02	14	1.679
S008_3	3;5.	male	146	430	11	0.026	0	0	1	0.002	22	0.051	38	9	0.021	14	1.677
S010_i	3;5.	male	63	129	0	0	3	0.023	1	0.008	3	0.023	5	2	0.016	2	1.4

AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS
% Phrase r	# Word r	% Word r	# Phrase r	% Phrase r	# Pauses	% Pauses	# Filled pauses	% Filled pauses	# SLD	% SLD	# TD	% TD	# Total (SLD+TD)	% Total (SLD+TD)	Weighted SLD
0	0	0	1	0.003	0	0	5	0.015	8	0.024	6	0.018	14	0.041	4.425
0.015	6	0.004	16	0.011	2	0.001	91	0.064	170	0.119	136	0.095	306	0.214	13.478
0	1	0.005	0	0	1	0.005	2	0.01	6	0.03	4	0.02	10	0.05	4.523
0.002	1	0.001	7	0.008	2	0.002	24	0.029	61	0.072	36	0.043	97	0.115	9.739
0.004	0	0	0	0	0	0	1	0.004	6	0.025	2	0.008	8	0.034	1.266
0.017	0	0	2	0.006	0	0	15	0.043	37	0.107	23	0.066	60	0.173	16.474
0.009	1	0.002	1	0.002	0	0	4	0.009	54	0.126	10	0.023	64	0.149	17.674
0	0	0	1	0.008	8	0.062	6	0.047	10	0.078	15	0.116	25	0.194	6.977

List of measures in FluCalc:

Each value is reported in **raw counts and proportions** (over **words OR syllables**)

- Total utterances in the sample
- Total intended words, as identified by MOR
- # Prolongation: raw count of sound prolongations
- # Broken word
- # Block
- # PWR (Part-word repetition)
- # PWR RU (Repetition units): iterations, or number of excess repetitions in a part-word repetition.
- # Phonological fragment –abandoned word attempts, e.g. *&+fr- tadpole*, where the speaker appears to change word choices
- # WWR (whole word repetition)
- # WWR RU (repetition units; please see PWR above)

FluCalc output, continued

- **# Phrase repetitions**
- **# Word revisions**
- **# Phrase revisions**
- **# Pauses** (hesitations)
- **# Pause duration** (if specified by coder)
- **# Filled pauses**
- **# SLD (stutter-like disfluency)**; summing over categories in columns Prolongations through whole-word repetitions (WWR), with the exception of columns reporting repetition units (RUs)
- **# TD (typical disfluencies)**: summing categories in columns labeled Phrase Repetitions through Filled Pauses)
- **# Total (SLD+TD)**: this sums all forms of disfluency, both stutter-like and typical, seen in the sample

Most critically, a reference value to distinguish normal disfluency from stuttering

- **Weighted SLD**. This is an adapted version of the SLD formula for distinguishing between typical disfluency and stuttering profiles in young children. It was originated by Yairi & Ambrose (1999, 2005).
- This formula multiplies the SUM of part-word and whole-word repetitions by the MEAN of the observed repetition units in the sample; it then adds this value to TWICE the sum of prolongations and blocks.
  - **A weighted score greater than 4.0** is considered greater than values obtained from typically fluent children and merits concern.

Fast clinical message:

- CUTTING TO THE CHASE:
- If you record and transcribe a single language sample
  - You can obtain about 40 measures of language use
  - Compare this profile to a very large database by gender and age
  - Obtain a profile of typical and atypical disfluencies
  - Compare the fluency profile to a well-respected formula to distinguish children who stutter (CWS) from other populations
- You will also have a permanent media record linked to the transcript to fully appreciate clinical progress.

If you run into problems and have questions when you try to do this yourself …

- Subscribe to the **chibolts** Google Group and post your question. It will be answered quickly (within 24 hours), usually by one of us or the CLAN programmer, Leonid Spektor.
- To subscribe, go to chibolts@googlegroups.com and provide a one-sentence statement about your research interests.
- The TalkBank webpage -- https://talkbank.org/ -- has a link called Google Groups where you can find this information.

Full Summary:

- These new free programs can help you in the following ways:
  - Easier, faster transcription of your sessions
  - Linkage to audio or video
  - Automatic and accurate linguistic analysis
  - > 40 language sample analyses in less than a minute
    - Norm-referenced over hundreds of children's data
  - The same sample can provide dozens of fluency analyses, including an evidence-based cut-off for distinguishing stuttering from other disfluency.
- It is all FREE and supported 24/7
- We invite your feedback to engage in continuous improvement to best support clinical assessment.

Questions?