PERSPECTIVES ▪

SIG 1

Tutorial

Assessment and Therapy Goal Planning Using Free Computerized Language Analysis Software

Nan Bernstein Ratner[a] iD and **Brian MacWhinney**[b] iD

[a] University of Maryland, College Park [b] Carnegie-Mellon University, Pittsburgh, PA

A R T I C L E   I N F O

A B S T R A C T

Background: We discuss a free software system (Computerized Language Analysis [CLAN]) that can enable fast, thorough, and informative language sample analysis (LSA).
Method: We describe methods for eliciting, transcribing, analyzing, and interpreting language samples. Using a hypothetical child speaker, we illustrate use KidEval to generate a diagnostic report.
Results: Given LSA results suggestive of expressive language delay, we analyze further using CLAN's Developmental Sentence Score and Index of Productive Syntax routines, and outline the child's use of Brown's morphemes.
Discussion: This tutorial provides an introduction to the use of free CLAN software. We discuss how LSA results can be used to structure therapy goals that address specific aspects of grammatical structure that the child may not yet demonstrate in their spoken language. Finally, we provide answers to common questions, including user support.

In this tutorial, we will discuss how to gather and analyze language samples for a broad range of age groups and clinical disorder types. These include evaluation of child language, adult acquired or degenerative language disorder, and fluency assessment. Given the focus of the special issue, we will concentrate primarily on preschool language sample analysis (LSA; child ages 2–6 years). We will discuss how to gather the most informative sample and how to obtain and use free, open access software that can perform in-depth analysis for both diagnostic and goal setting purposes. This software, called CLAN (**Co**mputerized **L**anguage **An**alysis), has been continuously supported by federal funding and development since 1986 (MacWhinney, 2014).

What Is CLAN?

The History of TalkBank/Codes for Human Analysis of Transcripts/CLAN

In 1984, Catherine Snow and Brian MacWhinney received support from the MacArthur Foundation to begin the Child Language Data Exchange System (CHILDES) project with a team of advisors and collaborators. The goal was to collect child language data from a large number of children and languages to help us better understand the process of language learning. In the 1980s, we developed a transcript coding system called CHAT (Codes for Human Analysis of Transcripts) that allowed us to rework data from existing corpora into a single consistent format. Once that format was established, we were able to begin the creation of the CLAN programs to process data in that format. These programs are able to analyze transcripts automatically for morphological and grammatical structure. This then further facilitated the development of automatic calculation of numerous child language measures. With the growth of computer technology, we were eventually able to make the programs and data available over the web. To

date, thousands of published articles have made use of CHILDES data or programs, as can be seen from a Google Scholar search.

In summary, the TalkBank initiative that hosts CLAN software consists of three major conceptual components: a massive archive of curated language samples, conventions for formatting transcripts (CHAT), and the CLAN software that can be used to analyze CHAT-formatted transcripts. TalkBank archives are in turn divided into separate Banks/repositories, primarily by clinical focus: child language (the CLDES), aphasia (Aphasia-Bank), fluency (FluencyBank), and others. The primary clinical function of these archives are to provide reference data ("norms") for measures that CLAN derives.

We find that some people think that CLAN itself is a language analysis measure. For instance, Finestack and Satterlund (2018) asked clinicians to check off diagnostic measures that they employ with children, and listed CLAN (and Systematic Analysis of Language Transcripts [SALT]) alongside traditional LSA measures such as mean length of utterance (MLU) and type–token ratio (TTR). CLAN is a tool that one can use to derive measures such as MLU and many other LSA outcomes, in the same sense that one can use either SPSS or Excel compute a t test.

CLAN is freely available for macOS, Windows, and Unix. It includes powerful linguistic analysis routines that can automatically supplement a basic transcript with accurate morphological and syntactic annotations, as we will illustrate. This, in turn, enables fast, accurate computation of important LSA metrics, such as MLU, DSS (Developmental Sentence Score), the IPSyn (Index of Productive Syntax), and others, which will be discussed in greater detail later in this tutorial. In addition, CLAN links transcripts directly to either audio or video recordings, providing faster and more accurate transcription, and allowing clinicians to monitor what clients actually sound like as they review transcripts for assessment and progress monitoring.

This tutorial is intended as a brief introduction to CLAN. After reading this *Perspectives* introduction, you may wish to next look at the "quick start" *Speech Language Pathology (SLP) Guide to CLAN,* which is freely available at https://talkbank.org/manuals/Clin-CLAN.pdf. This short (approximately 30 pages) guide is meant to walk users through all steps in the use of CLAN for clinical purposes, from downloading and installation of software, to transcript creation, through LSA. For more extensive documentation, you may wish to consult the CHAT transcription format manual at https://www.talkbank.org/manuals/CHAT.pdf, and the CLAN programs manual at https://www.talkbank.org/manuals/CLAN.pdf. All are updated on an almost daily basis, to be responsive to

changes in operating systems and continuous addition of new computing options. We also have a very extensive library of short video tutorials organized conceptually to illustrate steps in transcription and analysis at https://talkbank.org/screencasts/. A subset of five videos that can be viewed in their entirety in about an hour, meant to provide streamlined guidance for SLP users, is available at https://talkbank.org/screencasts/0SLP/. Our goal here is to show you some possibilities for making clinical work faster and more informative through use of these computer-assisted methods, so that you may explore their potential on your own. The CLAN system, which is part of the multinational, multilingual TalkBank initiative, also has full user support and user group listservs for those with questions. We address how to access these resources later in this tutorial, in the section on "Common User Questions."

Elicitation

A first consideration in obtaining a sample is whether you want to compare your client's performance to others for diagnostic purposes. For preschool children, the client group we focus on in this tutorial, the most typical context for a spoken language sample is adult–child free-play. Conventionally, the wisdom has been to use as many props or prompts as seem necessary to elicit a long enough sample to conduct typical linguistic analyses, which typically specify a minimum number of eligible utterances. For instance, Miller and colleagues note that, "the play context can be adjusted to meet individual preferences, individual interests, gender, culture, and experience" in the service of the largest possible sample (Miller et al., 2015, pp. 13–14). This enables child-centered, authentic data collection.

How long should a sample be? Because CLAN software derives a variety of independent LSA measures, length of a sample depends on which LSA measures the clinician considers important, for diagnosis, treatment planning, or both. Again, this can be compared to statistics: Some types of analyses will require a minimum set of data points to be valid. Many child LSA metrics, such as MLU (Brown, 1973), DSS (Lee, 1974), and the IPSyn (Scarborough, 1990) require at least 50–100 eligible utterances for computation. To be faithful to each of these protocols, some utterances cannot be used if they are partially unintelligible, are repeated, or do not have an identifiable subject or verb.

We note that for CLAN or any other software that derives these traditional measures of child language development, these requirements can make obtaining a language sample of sufficient size for analysis problematic with very young children, especially those with language delays or

pragmatic impairments that limit their social interaction with others, and this can add both transcription and analysis time. To address these concerns, newer adaptations of the IPSyn, which previously required 100 utterances, have been developed for computerized analysis of a sample size of 50 utterances (Yang, MacWhinney, & Bernstein Ratner, 2022). These adaptations reduce both data collection and transcription time considerably. For children in the preschool age range, free-play language profiles can also be referenced against a large corpus of data from almost 2,000 children ages 2–6 years in the TalkBank CHILDES archive, using the CLAN KidEval utility that we will discuss shortly, in the section on "Analysis."

Once the clinician has decided what information is desired from an LSA, they need to ensure that they collect a sample long enough for valid measurement. Historically, this has meant 50 unique, nonrepetitive, fully intelligible utterances for most measures, going all the way back to Mildred Templin's (Templin, 1957) guidelines, later extended by Roger Brown (1973). We find it useful to make simple tick marks on a sheet of paper as we hear the child produce such utterances, to ensure that we do not end the sample without sufficient data. Additionally, long experience in working with preschoolers has persuaded us to use certain toys and activities rather than others. For example, we no longer put trains or airplanes in the playroom, since they tend to elicit lots of "zoom, zoom, zoom" but less conversational language. We prefer play food and other interactive toys (a veterinarian office, dolls with many outfits, building materials). We sometimes leave things out of sets (missing slice of pizza, blocks in more than one container), or place a band aid on the baby doll or take away one of its shoes in order to try to observe questions and negatives.

For older, school-age children, there are also standard prompts to elicit language samples, including wordless picture books. The most widely used of these prompts are used by the SALT program and its associated database. This LSA software is discussed elsewhere in this special issue.

As we will discuss later, there can be two distinct reasons to conduct an LSA. One can be diagnostic, as discussed above. However, the second can be to establish relevant therapy planning goals and to measure therapy progress. In these cases, the language sample elicitation task can be chosen for its ecological validity in moving the client toward better communication in everyday settings. For children who will be expected to be able to narrate a story or explain how to play a game, the elicitation task can be set to resemble those needs. In each case, measuring what the client can currently do and what they may need to learn to do to improve communicative effectiveness is its own self-contained problem space.

Whatever the goal, it is best to record samples for later analysis. CLAN utilities allow linkage of audio or video media to the transcript, speeding transcription time and accuracy. Samples of linked transcripts can be seen at the CHILDES archive in TalkBank, such as https://sla.talkbank.org/TBB/phon/Eng-NA/Providence/Lily. A linked transcript can also be used to evaluate the client's behaviors holistically, including interpersonal, prosodic, articulatory, and fluency features of expressive behavior to be maintained over time. Finally, although some clinicians prefer to write down client speech in "real time," this may have the unintended impact of noting the most conspicuous utterances (those that are either longest, most complex, or inversely, those that contain obvious errors, etc.), thus distorting the overall picture of the client's typical communication profile (Berko Gleason & Phillips, in press).

Transcription

For the rest of this tutorial, we will assume that we have used the kinds of props described above to elicit a language sample from a boy age 3;6 (years;months; 42 months) who was observed playing with his mother. The parents have concerns about his pace of expressive language development. We video recorded about a 20-min sample, during which we tallied at least 50 intelligible, nonechoic, nonrepetitive utterances, which is the minimum for most traditional LSA measures, and then let the parent and child finish an activity before ending the interaction and moving on to other assessment targets, such as administration of a standardized expressive vocabulary test. So, now what do we do with this recording?

CLAN has several unique features that greatly ease the burden of language sample transcription. These include the ability to replay small segments of the audio or video for faster transcription. Additionally, CLAN's powerful English grammatical parser enables users to simply write what they hear using conventional orthography, including contractions; the program can then automatically identify morphemes, parts of speech, and sentence components for English and a dozen other languages. For English, the parser was recently evaluated against hand coding and judged to be above 95% accurate (MacWhinney et al., 2020). This can be compared against rare evaluations of the accuracy of SLP grammatical labeling, which Justice and Ezell (1999) found to be considerably lower, at about 70%.

Let us start with the process of linking the media (which can be any form of audio or any form of video other than DVD) to a transcript. The CLAN Transcriber Mode function can be used to make a first pass through the media. Each time the user presses the space bar, the editor enters a blank line for later transcription with time

marks for the beginning and end of the segment. Then, CLAN's Walker Controller function can be used to replay these segments (called "bullets") as many times as needed, while the clinician types what they hear. Examples of how this is done can be found at https://talkbank.org/screencasts/0SLP/Transcribing.mp4 and is also documented in the *SLP Guide to CLAN* at https://talkbank.org/manuals/Clin-CLAN.pdf.

Transcription conventions are quite simple. Files must start with an @Begin command and end with @End. There are a few required "headers" to specify the person/people talking in the file, and which media is linked, along with what language is being transcribed (because CLAN works in numerous languages). Figure 1 shows an example of a very short file, with its interface to mark speaker information.

The top few lines are automatically inserted after you insert the basic information about each speaker's name, age, gender, and group using the ID headers

command on the pull-down menu for TIERS, which is what CLAN calls speaker lines. Each line starts with a three-letter code to indicate the speaker. We prefer to use the abbreviation *CHI for all children, or *ADU for all adults, and leave the actual identifying information in the file header. This enables a clinician to analyze many transcripts at once, as we will show later in this tutorial.

The speaker identifiers are followed by a TAB and then the utterance. Lines are typed in basic English orthography, and most of the editor features are similar to those used in Microsoft Word. However, there is no need to capitalize the first words of utterances, because the use of caps is reserved for proper names/modifiers and the word "I" (and the grammatical analyzer uses this information in parsing what is typed).

The format of a transcript will look familiar to users of SALT, with minor differences. Each utterance is placed on a line, using conventional definitions; in our work (see page 13 of the *SLP Guide to CLAN*), we use "two out of three"

Figure 1. A short sample Codes for Human Analysis of Transcripts transcript (before grammatical annotation). Adapted from the POLER corpus at the Child Language Data Exchange System TalkBank site (https://childes.talkbank.org/access/Clinical-MOR/POLER.html).
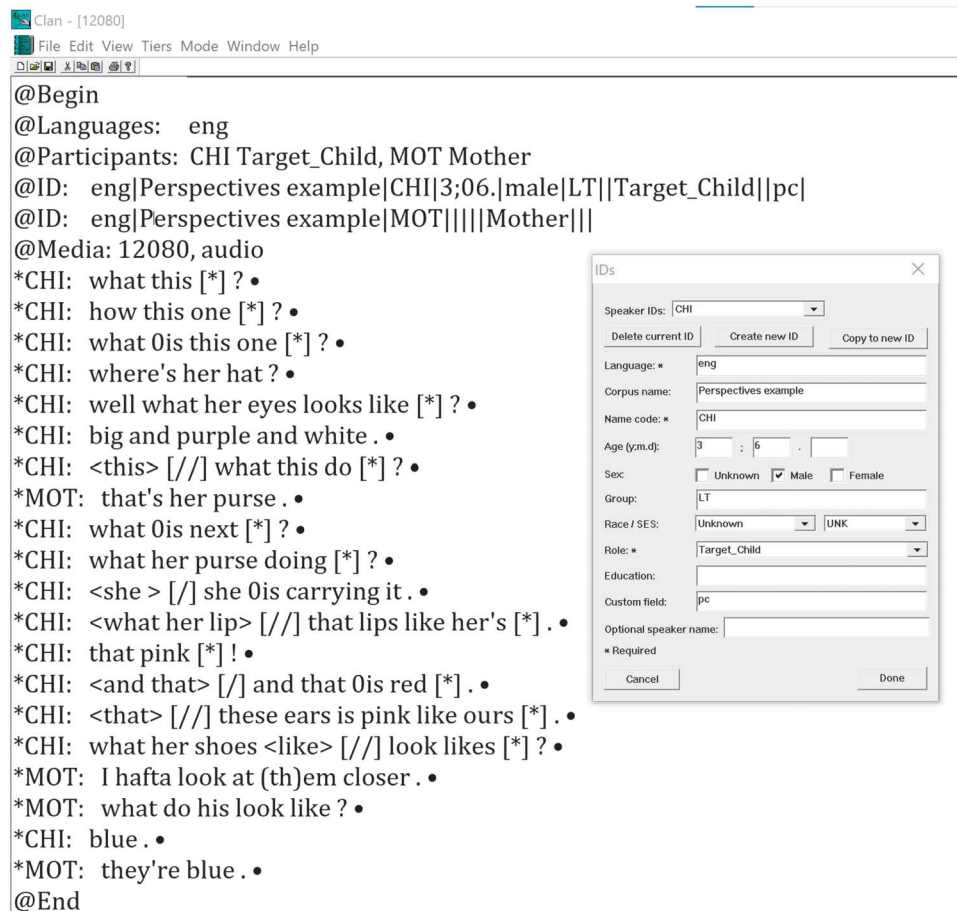
criteria to demarcate the ends of utterances and beginnings of new utterances: silence, grammatical structure, and terminal intonation contour. Each utterance must end in a punctuation mark, such as a period, a question mark, or a small set of marks for utterances that are interrupted or trail off, as listed in the SLP and CHAT guides. You can use contractions (*I'm, we're, don't,* etc.), and you do not need to mark off morphemes, as is required by some other systems, such as SALT. In other words, CLAN lets you simply write *dogs, played, was,* and *doesn't,* while alternative LSA systems such as SALT (Miller et al., 2015) or SUGAR (Pavelko & Owens, 2017) require you to separate morphemes. For instance, SALT would use *dog/s* for the plural, or *play/ed* for the past tense.

However, our first short file does have some symbols that require some explanation and we have provided a guide to them in Table 1. For instance, we see < > around some words, and / inserted between some phrases. These are to mark repeated materials that we would not necessarily want to include in an MLU count or count of total number of words (TNW) because they are repetitions, restarts, or what some authors have called "mazes." We also have marked some words with a leading zero (e.g., 0is) to show that a word was expected but not produced. Likewise, CHAT transcription asks the clinician to mark normal disfluencies with the &- marker to exclude them from MLU, TNW, and TTR (consider, for instance, the difference between "well you know" as a series of filled pauses signifying uncertainty, vs. an expression such as "you know him well."). There is an additional small set of marks, some of which we see in Figure 1, that can be used to improve the accuracy of LSA measures, as shown in Table 1.

Because you can automatically replay the media from the segments you entered using Transcriber Mode, and because you do not need to manually identify any aspects of the grammar while you transcribe, CLAN transcripts can be completed very quickly. For example, you can instruct the system to "loop" each short segment 3

times while you type, all without leaving the transcription screen. Because of this function, in our extensive work over the past few decades, we estimate that CLAN transcripts take about 1/3 of the time of conventional transcripts made by simply viewing a video or listening to audio through a media player. Thus, we estimate that most transcripts for clinical work can be made at the rate of about 5× actual sample length (e.g., a 15-min sample can be transcribed in a little more than an hour if the clinician is only interested in language analyses, as opposed to greater detail as might be required by a subsequent fluency analysis of the same sample, or more detailed research questions that require annotation or coding). For that hour investment in time, the multiple LSA measures we describe are then produced in seconds.

Analysis

There are two steps to analyzing a CLAN transcript. The first is to insert grammatical information into your "starter" transcript, and the second is to generate a clinical profile. Clinical profiles can be generated for child and adult language (KidEval/Eval), and for child or adult fluency features (FluCalc). Here, we concentrate on child LSA using KidEval, one of many packaged routines in CLAN that combines a number of individual traditional LSA analyses (more than three dozen) using a single command (much like the SALT report function).

### Inserting Grammatical Information Into Your Transcript

Unlike other LSA systems, CLAN can automatically annotate your transcript for part of speech, morphological composition, and grammatical relationships. As a side effect, it can check your transcript for spelling errors that can interfere with this process. We have an open-access tutorial to walk you through these steps at https://talkbank.org/screencasts/0SLP/MOR&IPSyn.mp4. The first step is to download the grammar for the language you are working in. Although CLAN provides analyzers for 12 languages, we will assume here that you are analyzing transcripts in English. Using the FILE option > get MOR grammar, we would select English (you must be connected to the Internet to get the grammar and install it for the first time, but after that, you do not need to use Internet to work with CLAN). You will know that the grammar has been installed correctly if your CLAN commands window shows the MOR library (mor lib) ending in ENG for English, as Figure 2 illustrates.

Once the English grammar is installed, the user simply types the word MOR into the Commands window,

Table 1. Examples of Computerized Language Analysis codes for marking words and utterances.

Symbol	Use	Example
. (period)	Utterance delimiter	I have it.
? (question mark)	Utterance delimiter	Is it mine?
+…	Utterance trails off	I see two +…
[*]	Error	He's gots [*] two.
&-	Filler	&-um
0word	Obligatory word deleted	0is he going? (gloss: he going?)
<word> [/]	Repeated material; maze without correction	<he> [/] got it.
<words> [//]	Revise material; maze with correction	<he made> [//] he cooked dinner.

Figure 2. How to verify English grammar installation. The ending letters (eng) tell the user that the English grammar has been installed and is ready to use. It needs to only be installed once per machine, although some users update the grammar periodically, since it continues to add entries for new words.
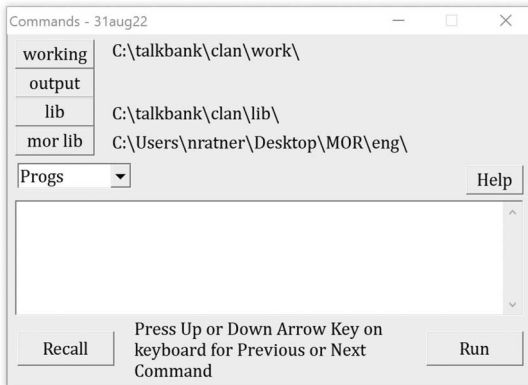
inputs the transcript file, and the program uses extensive dictionary and syntactic resources to annotate the original file. Figure 3 below shows how a basic transcript is automatically annotated for English part-of-speech and grammatical relationships. Once completed, a simple basic transcript (left) now contains new information (right) necessary to conduct most LSA analyses (MLU, DSS, IPSyn, and various lexical diversity analyses).

Figure 3. Adding grammatical information to a transcript. Left: a basic transcript. Right: portions of the same transcript after running MOR to add grammatical and syntactic information to the basic file.

```
@Begin
@Languages:   eng
@Participants: CHI Target_Child, MOT Mother
@ID:   eng|Perspectives example|CHI|3;06.|male|LT||Target_Child||pc|
@ID:   eng|Perspectives example|MOT|||||Mother|||
@Media: 12080, audio
*CHI:  what this [*] ? •
%mor:  pro:int|what det:dem|this ?
%gra:  1|2|LINK 2|0|INCROOT 3|2|PUNCT
*CHI:  how this one [*] ? •
*CHI:  what 0is this one [*] ? •
%mor:  pro:int|what 0cop|be&3S det:dem|this pro:indef|one ?
%gra:  1|2|SUBJ 2|0|ROOT 3|4|DET 4|2|PRED 5|2|PUNCT
*CHI:  where's her hat ? •
%mor:  pro:int|where~cop|be&3S det:poss|her n|hat ?
%gra:  1|2|SUBJ 2|0|ROOT 3|4|DET 4|2|PRED 5|2|PUNCT
*CHI:  well what her eyes looks like [*] ? •
%mor:  co|well pro:int|what det:poss|her n|eye-PL v|look-3S prep|like ?
%gra:  1|5|COM 2|5|LINK 3|4|DET 4|5|SUBJ 5|0|ROOT 6|5|JCT 7|5|PUNCT
%gra:  1|2|SUBJ 2|0|ROOT 3|2|OBJ 4|2|PUNCT
*CHI:  big and purple and white . •
%mor:  adj|big coord|and adj|purple coord|and adj|white .
%gra:  1|0|INCROOT 2|1|CONJ 3|2|COORD 4|1|CONJ 5|4|COORD 6|1|PUNCT
*CHI:  <this> [//] what this do [*] ? •
%mor:  pro:int|what det:dem|this v|do ?
%gra:  1|3|LINK 2|3|DET 3|0|ROOT 4|3|PUNCT
*MOT:  that's her purse . •
%mor:  pro:dem|that~cop|be&3S det:poss|her n|purse .
31aug22[E|CHAT] * 6
```

The new lines that were automatically added include a tier labeled %mor (for morphology) that labels each item in the transcript for part of speech (such as noun, pronoun, etc.). The second added tier is labeled %gra (for grammar) that identifies the functional categories, such as subject, predicate, object, and so forth. These labels are used in some traditional LSA measures, such as the IPSyn. As we noted previously, although the process of labeling these items by hand would be time-consuming and painstaking (Long, 2001) and subject to clinician error (Justice & Ezell, 1999), the CLAN parser for English has been appraised at approximately 95%–97% accuracy (MacWhinney, 2008; MacWhinney et al., 2020).

Conducting the LSA

Once a file has been annotated for morphology and grammar, numerous analyses can be performed on a single file within seconds, or on a large set of files (e.g., progress reporting for your caseload) at the same time. Although individual LSA analyses (such as MLU, TTR) can be run one at a time, our favorite program is KidEval, akin to the SALT report, that analyzes a file for more than three dozen features at the same time. Its spreadsheet format also references preschool children's performance (in free-play/conversation) against that of hundreds of same-sex, same-age children in the CHILDES database. Full demographics (age, diagnosis, and primary study citation) are described in detail for 1,454 typical and language-impaired/delayed children ages 2–6 years in the works of Yang et al. (2021) and Yang, MacWhinney, and Bernstein Ratner (2022), and Yang, Rosvold, and Bernstein Ratner (2020). Because not all children provided enough eligible utterances for every LSA computation done in KidEval, the actual number of comparison files for any given child will vary slightly, and the user can also decide whether or not to further narrow comparisons by sex. Reference values for older children and for elicited narratives are pending, but the program will still provide values for all measures that can be used to measure a client's progress over time. The video that demonstrates how to automatically insert grammatical information that we referenced above, which is only about 8 min long, shows how easy and fast it is to run KidEval.

Interpretation

A sample report is shown below in Figure 4; we have "wrapped" the columns for easier reading and deleted some columns that are not necessary for this discussion. From the first column on the left through the end of the report, the name of a measure is given in the top row; the child's value is in the second row. Whether this score (for preschool children [ages 2–6 years]) is within the standard deviation for age (and sex, if selected), is in Line

Figure 4. A sample KidEval report. We have "wrapped" columns for easier reading, as well as added shading. The actual report is an Excel file.

	A	E	F	P	Q	R	S
1	File_DB	Age(Month)	Sex	Total_Utts	MLU_Utts	MLU_Words	MLU_Morphemes
2	42pc\12080.cha	42	male	223	222	2.504	2.77
3	+/−SD	.	.	0.018	0.087	−1.09	−1.15
4						*	*
5	Mean Database	.	.	220.283	209.083	3.614	4.113
6	SD Database	.	.	152.172	149.092	1.04	1.192
7	Number files	.	.	180	180	180	180
8							
9	PLEASE SEE SLP MANUAL FOR RECOMMENDED INTERPRETATION.						
10							
11	+/− SD * = 1 SD, ** = 2 SD						
12	Database keywords: 3;6-3;11						
13	Database date: 2022-01-28, 12:49						

	W	X	Y	Z	AA	AB	AC
1	FREQ_types	FREQ_tokens	FREQ_TTR	NDW_100	VOCD_D_optimum_average	Verbs_Utt	Word_Errors
2	149	674	0.221	51	47.83	0.368	79
3	−0.489	−0.298	−0.498	−0.02	−0.028	−1.538	3.471
4					*	**	
5	187.267	870.483	0.258	51.111	48.113	0.707	4.467
6	78.312	659.453	0.075	5.697	9.958	0.221	21.475
7	180	180	180	180	180	180	180

	AE	AF	AG	AH	AI	AJ
1	retracing	repetition	DSS_Utts	DSS	IPSyn_Utts	IPSyn_Total
2	18	19	50	6.38	50	35
3	2.664	1.928		−1.159		−1.434
4	**	*		*		*
5	3.578	5.667	50	8.989	50	50.224
6	5.415	6.915	NA	2.252		10.616
7	180	180	138	138	174	174

	AL	AM	AN	AO	AP	AQ	AR	AS
1	*-PRESP	in	on	*-PL	irr-PAST	*-POSS	u-cop	det:art
2	13	4	0	33	1	1	14	43
3								
4								
5								

	AT	AU	AV	AW	AX	AY	AZ
1	*-PAST	*-3S	irr-3S	u-aux	c-cop	c-aux	Total_non_zero_mor
2	1	14	4	4	13	1	13

3. The mean for the reference children is in Line 4 (with its standard deviation in parentheses), and finally, the number of comparison children from the reference database is listed in the last row.

In this case, we see that the child, a boy aged 3;6, has produced 223 utterances, of which all but one meet Brown's criteria for computing MLU. He's certainly talkative, but what CLAN KidEval wants to know is the nature of those utterances—what grammar is evident, and which does not appear. Absent structures can then be probed to see if the child knows how to use them, but does not find a need in this context. However, most KidEval measures are less concerned with quantity indices (e.g., TNW) than age-appropriateness of structure.

Utterance Length and Grammatical Measures for Diagnostic Purposes

This child's MLU in words and morphemes are 2.5 and 2.77, respectively, far below that expected for his cohort, which had values of 3.6 and 4.1. Thus, the standard deviation row contains an asterisk to flag that these values are more than 1 SD (−1.09, −1.15) from that expected of boys his age.

Lexical measures (TTR, NDW, and vocabulary diversity) are listed in the next few columns. We will not spend much time in this tutorial on these LSA values, because our recent research (Yang, Rosvold, & Bernstein Ratner, 2022) has produced some serious reservations over use of these measures to diagnose child language skills. TTR does not show a growth function over childhood, neither TTR nor NDW distinguish delayed from typical children, and all three measures are statistically impacted by the activities and props used during the language sample elicitation process.

Verbs per utterance, a proxy for clausal embedding, is in the next column. Once again, this child's value of .37 is more than 1.5 *SD*s from the expected mean; only slightly more than one third of the child's utterances contained a verb. Notice that not all typical children's utterances in conversation contain verbs. For example, if an adult asks the child, "What do you want to play with?" an appropriate response is simply "blocks," not "I want to play with blocks." However, as a proxy for the complexity of a child's spoken language, verb density can be a useful measure. This child's use falls far below that seen in typical children his age.

Grammatical Measures for Both Diagnostic and Therapy Planning Purposes

We will concentrate on the next few measures, DSS (Lee, 1974; Lee & Canter, 1971), the IPSyn (Altenberg et al., 2018; Scarborough, 1990), and Brown's (Brown, 1973) original list of 14 grammatical morphemes critical to English language development. All three can inform assessment and have good potential for structuring therapy intervention goals. All are typically described in major texts on child language disorders, such as Paul et al. (2018).

Describing DSS and the IPSyn in More Detail

A very cursory distinction between DSS and the IPSyn is that DSS measures use of specific words from eight parts of speech (e.g., conjunctions, indefinite pronouns, quantifiers) that increase gradually in developmental difficulty and mark development of more sophisticated sentence structure. For instance, a negative form such as *no* is less developmentally sophisticated than *don't,* and both are certainly less sophisticated than *neither*. In this vein, individual words in the transcript are identified from Lee's 1974 guide and given point scores from 1 to 8 (least to most advanced). In contrast, the IPSyn examines what we might recall from linguistics classes as "phrase structure" or rewrite rules. In this case, development of noun phrases (NPs) might be tracked by giving credit (up to 2

points for two examples) of forms such as a single noun (*Mommy, I*). A next development might be article + noun NPs, such as *the baby* or *a truck*. Next, the child might use modifiers: *a big teddy* or *the yellow truck*, and so on. The IPSyn tracks structures in four categories: noun phrases, verb phrases, questions/negatives, and sentence structures. Our recent work (Yang, MacWhinney, & Bernstein Ratner, 2022) has shown statistically significant ability of the verb phrase and sentence structure scales to discriminate younger from older children, and typical from delayed language. Both IPSyn and DSS analyze question and negative constructions; however, not all sampling interactions provide contexts for appropriate use of questions or negatives. We recommend that these forms be probed in a more deliberate fashion, and not from an unstructured play session. While DSS and the IPSyn can be used in initial assessment, our favorite use of them is in goal setting, as we will discuss shortly in the section on *Using DSS and the IPSyn to Set Therapy Goals for Children*.

Both DSS and the IPSyn are rather daunting to compute by hand, both in terms of time and accuracy (see Hughes et al., 1994, and Altenberg & Roberts, 2016, for a relevant discussion), but CLAN finds such algorithms rather easy and calculates them extremely accurately in seconds (MacWhinney et al., 2020). Both also require that only some of the child's utterances be analyzed and others excluded, an additional headache for many clinicians. For instance, both measures exclude utterances with unintelligible sections, and both require a minimum of 50 eligible utterances. DSS requires that an utterance minimally contain a noun and a verb. The IPSyn only counts the first two examples of individual syntactic constructions. Thus, columns such as DSS utts (utterances) and IPSyn utts on the KidEval printout first list how many of the child's utterances were eligible for analysis; the next column then provides the score (with the same references to CHILDES cohort values as in earlier columns). In each case, this child's total score is well below what is expected; for DSS, the score is more than 1 *SD* below that expected for age. For the IPSyn, it is almost 1.5 *SD*s below.

Similarly, tracking use of Brown's grammatical morphemes can be informative both in assessment and in goal setting. Some definitions are typically useful in interpreting the KidEval output. For instance, u-cop, u-aux, c-cop, and c-aux stand for uncontractible/contractible copula/auxiliary. The copula is the main verb to be in English ("he's nice,"), whereas the aux is the "helping verb" ("he's going). Notice that in both cases, we contracted, because we can say both "he is nice" and "he's nice." Thus, these are contractible contexts. In contrast, when asking and answering questions ("*is* he nice/going?" "Yes, he *is*"), we

cannot contract (uncontractible context). We can't say, *"'s he nice; *yes, he's." Interestingly, children often learn these uncontractible forms first, because they are rather conspicuous, and they hear them often ("Who's a good boy? You are!!"). So a child may indeed say, "Yes, I am" before including the cop/aux in "He_nice/going." Similarly, a therapist may find it easier to teach such forms, if missing, because of their salience (Fey et al., 2003).

The forms 3S and irr3S stand for the regular third-person singular present form and its irregular versions in Mainstream American English (MAE; "I run, you run, we run, they run, he/she/it runs"). A few select but common verbs in MAE are irregular—we say *has* rather than *haves* and *does* rather than do-s [duz]; note the vowel change in *does,* unlike the more regular form *goes,* which is pronounced as you might expect it to be (*I go, she goes*).

Regardless of problems in language use that are seen throughout the KidEval report, the child does seem to use most of Brown's morphemes at some point in the play session. The final column shows that 13 of the 14 were observed at least once.

Before leaving this section, please note that we have introduced the term *MAE* here. That is because varieties of English vary in syntactic structure. For example, African American English regularizes some third-person markers, or optionally produces copulas or auxiliary forms in contractible (but not uncontractible, environments). For this reason, it is important that any LSA work conducted with children who speak varieties of English other than MAE be adjusted to recognize these differing but acceptable forms. Most work on LSA in English has centered on MAE to the exclusion of other English varieties. We are currently working to address this disparity (e.g., a release of Black English Sentence Scoring ([Hyter, 1984] as part of CLAN software is planned for early 2023). For more discussion of this issue, please see the work of Overton et al. (2021) for further examples and suggestions.

Using DSS and the IPSyn to Set Therapy Goals for Children

If LSA or other assessment results suggest that a child should receive therapy, we can make use of DSS and the IPSyn for setting expressive language goals. To do this, we run each as separate routines (not just the summary score in the KidEval output) to get a detailed report. The command is just DSS (or IPSYN) + t*CHI (meaning analyze the child's utterances) and then the file name. We will give an abridged example from DSS to show why we value this kind of analysis. Note that we only provide part of the analysis to explain the concept; not all of the child's utterances are shown.

Consider the child above. His MLU, as computed by KidEval, was very short. Clearly, he needs to produce longer utterances. However, "say longer utterances" is not a very precise or informative goal. In the figure below (see Figure 5), we show a portion of this child's DSS analysis.

We can notice a few things. Most, but not all, of his DSS "credits" are from relatively low-numbered (less mature) forms (items listed as 1, 2, or 3). Some columns have few entries at all (such as indefinite pronouns and secondary verbs). In this case, we might decide to lengthen utterances by increasing the child's repertoire of structures in these categories.

The child asks numerous questions, but most are ungrammatical. In fact, the DSS score is lowered considerably by the fact that few utterances receive the sentence point for well-formed utterances. Because the child appears to be on the cusp of putting together adultlike questions (and appears to make good use of them pragmatically), questions could be a good target for intervention.

Similarly, if our IPSyn shows that most noun phrases are short, and lack modifiers or prepositional phrases, we could start teaching those as a way to expand utterance length. The use of both DSS and the IPSyn to structure therapy goals has been discussed in several recent American Speech-Language Hearing Association (ASHA) journal articles (Finestack et al., 2020; Garbarino et al., 2020; Pezold et al., 2020).

Figure 5. A portion of a Developmental Sentence Score analysis of a child with delayed expressive language. Low scoring categories or absent structures can serve as concepts for intervention planning and prioritization.


```
From file <c:\talkbank\clan\work
Developmental Sentence Analysis

    Sentence                            |IP |PP |MV |SV |NG |CNJ|IR |WHQ|S|TOT|
where's her hat ? ·                     |   |2  |1  |   |   |   |   |1 |2 |1| 7 |
well what her eyes looks like [*] ? ·   |   |2  |2  |   |   |8  |   |1 |2 |0| 15|
<this> [//] what this do [*] ? ·        |1  |   |1  |   |   |   |   |1 |2 |0| 5 |
what her purse doing [*] ? ·            |   |2  |   |4  |   |   |   |1 |2 |0| 9 |
<what her lip> [//] that lips like her'|1  |   |1  |   |   |8  |   |  |  |0| 10|
  s [*] . ·
a lip lives here . ·                    |   |   |2  |   |   |   |   |  |  |1| 3 |
I don't want to . ·                     |   |1  |3 1|   |4  |   |   |  |  |1| 10|
I just want to &+st with this [*] . ·   |1  |1  |1  |   |   |   |   |  |  |0| 3 |
<that> [//] these ears is pink like     |   |3 3|1  |   |   |   |   |  |  |0| 7 |
  ours [*] . ·
what her shoes <like> [//] look likes|  |2  |2  |   |   |   |   |1 |2 |0| 7 |
  [*] ? ·
+, that shoes them are [*] . ·          |1  |3  |2  |   |   |   |   |  |  |0| 6 |
where's somebody else ? ·               |3  |   |1  |   |   |   |   |1 |2 |1| 8 |
here's a farmer . ·                     |   |1  |1  |   |   |   |   |  |  |1| 2 |
where's the farm animal ? ·             |   |   |1  |   |   |   |   |1 |2 |1| 5 |
I'm gonna do it like this . ·           |   |11|1 |2 3|2  |   |   |6  |  |1| 18|
and this is (a)nother chicken . ·       |1  |3 |1  |1  |   |   |3  |  |  |1| 9 |
where's the baby pig ? ·                |   |   |1  |   |   |   |   |1 |2 |1| 5 |
and look a baby cow
and a mommy cow.                        |   |   |1  |   |   |   |3 3|  |  |1| 8 |
31aug22[F][TEXT] * 32
```

Additional Considerations in Using CLAN to Perform LSA

Use of CLAN With Other Groups of Clients

We have focused on child language in this tutorial, given its inclusion in a Special Interest Group (SIG) 1 *Perspectives* forum. However, CLAN utilities also include "packages" for the analysis of adult language (aphasia, dementia, mild cognitive impairment [MCI]) using EVAL, as well as fluency analysis for clients who stutter or clutter (FluCalc).

Gathering samples from adults with acquired language disorders: There are numerous standard protocols for obtaining short narratives from adults with neurogenic disorders. The AphasiaBank protocol uses descriptions of the "Cookie Theft" picture, procedural explanation of how to make a peanut butter sandwich, retellings of the Cinderella story, and personal narratives. A client's responses on these tasks can be compared with the large comparison reference data at AphasiaBank (aphasia.talkbank.org) that can be used to position the client's expressive language against those of other similar-age typically functioning adults and those with previously identified profiles of language impairment. For instance, EVAL produces reports that analyze the client's use of specific parts of speech, useful in assessing and setting therapy goals for individuals with agrammatic output. Exploratory work has now identified many LSA features that can be used to track the onset of MCI and dementia (see Faroqi-Shah et al., 2020, for a discussion).

Using CLAN with children and adults who stutter: To measure fluency of oral reading, the FluencyBank protocol uses standard passages, such as those used by the Stuttering Severity Instrument (Riley, 2009). When transcribing, CLAN provides standardized codes for disfluencies that can be used across language communities and orthographies (as we see in Figure 6). They are listed in the SLP and CHAT guides to CLAN, and symbols are inserted via an easy pull-down window.

Once a transcript has been coded using these symbols, CLAN's FluCalc command can generate a full profile of typical and stutter-like disfluencies (SLDs) in the client's speech, as well as compute the weighted disfluency score (weighted SLD; Ambrose & Yairi, 1999). This evidence-derived score has been shown to be crucial for distinguishing between typical, "developmental disfluency," and childhood stuttering. Having said that, it is quite complex to compute, since it requires taking the mean number of iterations in a client's repetitions, and penalizing blocks and prolongations with extra points, among other computations. FluCalc computes these measures and others in seconds. Moreover, for a child with

Figure 6. English parts of speech.

English Parts of Speech

Category	Code
Adjective	adj
Adjective-predicative	adj:pred
Adverb	adv
Adverb-temporal	adv:tem
Communicator	co
Complementizer	comp
Conjunction	conj
Coordinator	coord
Determiner Article	det:art
Determiner-demonstrative	det:dem
Determiner-interrogative	det:int
Determiner-numeral	det:num
Determiner-possessive	det:poss
Filler	fil
Infinitive	inf
Negative	neg
Noun	n
Noun-letter	n:let
Noun-plurale tantum	n:pt
Proper Noun	n:prop
Onomatopoeia	on
Particle	part
Postmodifier	post
Preposition	prep
Pronoun-demonstrative	pro:dem
Pronoun-existential	pro:exist
Pronoun-indefinite	pro:indef
Pronoun-interrogative	pro:int
Pronoun-object	pro:obj
Pronoun-personal	pro:per
Pronoun- possessive	pro:poss
Pronoun-reflexive	pro:refl
Pronoun-relative	pro:rel
Pronoun-subject	pro:sub
Quantifier	qn
Verb	v
Verb-auxiliary	aux
Verb-copula	cop
Verb-modal	mod

concomitant problems, the same transcript can be used to perform both a fluency and language analysis, a real saving of time and energy.

Tutorials more directly related to these kinds of analyses can be found in the *SLP Guide to CLAN,* referenced above, and tutorial screencasts at the TalkBank site. TalkBank even hosts programs for detailed analysis of phonological features of language using the Phon utility. All are free and have extensive user supports provided by U.S. federal funding initiatives.

Common User Questions About CLAN

As we have said, there are no costs associated with using CLAN or any of its resources (manuals, mailing lists, archival samples, etc.). However, we do encounter some questions often enough that we would like to address them as we come to the conclusion of this tutorial.

Why Do I Still Need to Transcribe My Language Sample?

Work to enable automatic speech recognition (ASR) of clinical samples is very much a work in progress. We have had success using the ASR provided by Amazon Web Services for fluent adult speakers. However, when we work with small children with developmental delays, and transcripts with long pause breaks or stuttering, our experience has been that recordings of this type are not yet well handled by these utilities. The same will be true with clients with articulation or phonological disorders. As a result, we spend more time trying to correct mistranscription of misarticulated child attempts than if we just did it by hand. However, companies and research groups are working actively to improve this situation. In a nice example of reciprocal benefits in research, to train and validate these new models, they are often relying on data from TalkBank. This represents another positive result of the open data-sharing of corpora in TalkBank.

Unlike SALT, which offers to facilitate transcription of language samples for a fee, TalkBank is not configured to vend services or products. It is always possible for a clinician or practice to work with a transcription service. In addition, CLAN can import SALT files to perform analyses that are available in CLAN but not in SALT, such as the IPSyn and DSS.

I Work With Older Children/Students—Can I Use CLAN to Analyze Written Language?

You can use KidEval, IPSyn, and DSS with students' written language samples to identify areas of intervention need, as we did in the example earlier. If a child has produced any form of keyboarded sample, the clinician need only save it as plain text (*.txt). Then, the TEXT2CHAT command in CLAN will turn the sample into a transcript that you can analyze in the same way that we described above. See the full CLAN manual and the *SLP Guide to CLAN* for more detailed instructions. Currently, TalkBank does not have reference values for older children's performance, although we are working to create them in the near future (MacWhinney & Bernstein Ratner, 2022).

I Tried Using CLAN and I Am Getting Error Messages/Am Lost/Am Frustrated

TalkBank hosts free subscription services via Google groups (see https://www.talkbank.org/share/email.html) that enable users to communicate their difficulties and get very timely responses (often within the hour). We suggest that you join a Google user group at TalkBank (such as CHIBolts, which is mainly for CLAN users, or the CHILDES user group, which has a primary interest in child language research and practice, similar to ASHA SIG 1). If you are not able to use CLAN without difficulty, simply write to the user group, attach the file you wrote and the error message you received, and other users or the TalkBank support staff will get back to you fairly quickly (you can also write to Nan Ratner at nratner@umd.edu directly for help). Our grant funding also enables us to run training sessions for groups of users (for instance, members of a practice, or an in-service for SLPs in school systems). These can be done very easily by teleconferencing, or in person. So do not let any possible early frustration slow you down. Our experience is that most problems are very simple to deal with and that clinicians are then amazed at how much easier it is to assess and track clients' expressive language skills.

Limitations of CLAN for LSA

As with most LSA methods, the first limitation of using CLAN or other software, or doing LSA by hand, is time: We do not currently have ASR that can faithfully transcribe what clients say, particularly young children with immature phonology, soft voices, and, in many cases, toy play clatter in the audio signal that real listeners can ignore but many ASR systems cannot handle. Having said this, people are working on the problem quite actively, and we hope that progress will be made in this area. It is the major time commitment in LSA if the clinician wants to measure more than simple quantitative measures, such as TNW, which conceivably can be done in real time by a trained assistant or observer or automated systems such as Language ENvironment Analysis (LENA). LENA can estimate quantity (but not describe content) using algorithms based on syllable counts in running speech.

A second limitation is the time required to learn how to transcribe and analyze. In our experience, this time is not extensive, and most of our local users are on their way within 1 hr of practice after watching our set of videos that take about an hour to watch from beginning to end. The system does have extensive feedback to users if it sees problems in transcript construction (using our CHECK function). In return, the user saves tremendous time in analysis, both for assessment and therapy planning purposes.

A third limitation might be called information overload. KidEval provides so much feedback about a child's performance that it requires the clinician to be selective in which measures are most important for what purposes with which clients. For instance, knowing which of Brown's morphemes is present or absent may be useful for a young toddler, while constructing targets for lengthening utterances might suggest the use of the IPSyn to catalogue constructions that are present, and others that are not observed. CLAN reports multiple measures of lexical diversity, in part because the field has yet to determine which are best for clinical purposes (see Yang et al., 2021). As with the results of a blood panel from your physician, some selectivity and interpretation on the part of the clinician is imperative in extrapolating from the quantitative output to decision making and recommendations.

Conclusions

As Hughes et al. (1994) pointed out a number of years ago, graduate students in SLP are often instructed in the rather arduous task of performing detailed LSA as part of their educational program. However, upon graduation, they find that they have little time (and diminished knowledge of some of the linguistic concepts) required to use these valuable tools in everyday work. Computers can help to make informative analysis of language samples both quick and more accurate. The TalkBank initiative and CLAN software aim to provide clinicians with free, easy to use, informative, and accurate LSA tools. We hope that you find them useful.

Acknowledgments

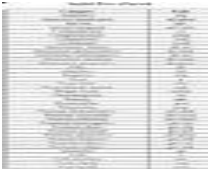## References

Altenberg, E. P., & Roberts, J. A. (2016). Promises and pitfalls of machine scoring of the index of productive syntax. *Clinical Linguistics & Phonetics, 30*(6), 433–448. https://doi.org/10.3109/02699206.2016.1139184

Altenberg, E. P., Roberts, J. A., & Scarborough, H. S. (2018). Young children's structure production: A revision of the index of productive syntax. *Language, Speech, and Hearing Services in Schools, 49*(4), 995–1008. https://doi.org/10.1044/2018_LSHSS-17-0092

Ambrose, N. G., & Yairi, E. (1999). Normative disfluency data for early childhood stuttering. *Journal of Speech, Language, and Hearing Research, 42*(4), 895–909. https://doi.org/10.1044/jslhr.4204.895

Berko Gleason, J. & Phillips, B. (in press). The development of language: An overview and a preview. In J. B. Gleason & N. B. Ratner (Eds.). *The development of language* (10th ed.). Plural.

Brown, R. (1973). *A first language: The early stages.* Harvard University Press. https://doi.org/10.4159/harvard.9780674732469

Faroqi-Shah, Y., Treanor, A., Ratner, N. B., Ficek, B., Webster, K., & Tsapkini, K. (2020). Using narratives in differential diagnosis of neurodegenerative syndromes. *Journal of Communication Disorders, 85,* 105994. https://doi.org/10.1016/j.jcomdis.2020.105994

Fey, M. E., Long, S. H., & Finestack, L. H. (2003). Ten principles of grammar facilitation for children with specific language impairments. *American Journal of Speech-Language Pathology, 12*(1), 3–15. https://doi.org/10.1044/1058-0360(2003/048)

Finestack, L. H., Rohwer, B., Hilliard, L., & Abbeduto, L. (2020). Using computerized language analysis to evaluate grammatical skills. *Language, Speech, and Hearing Services in Schools, 51*(2), 184–204. https://doi.org/10.1044/2019_LSHSS-19-00032

Finestack, L. H., & Satterlund, K. E. (2018). Current practice of child grammar intervention: A survey of speech-language pathologists. *American Journal of Speech-Language Pathology, 27*(4), 1329–1351. https://doi.org/10.1044/2018_AJSLP-17-0168

Garbarino, J., Bernstein Ratner, N., & MacWhinney, B. (2020). Use of computerized language analysis to assess child language. *Language, Speech, and Hearing Services in Schools, 51*(2), 504–506. https://doi.org/10.1044/2020_LSHSS-19-00118

Hughes, D. L., Fey, M. E., Kertoy, M. K., & Nelson, N. W. (1994). Computer-assisted instruction for learning developmental sentence scoring. *American Journal of Speech-Language Pathology, 3*(3), 89–95. https://doi.org/10.1044/1058-0360.0303.89

Hyter, Y. D. (1984). *Reliability and validity of the Black English sentence scoring system* [Unpublished master's thesis]. Western Michigan University. Retrieved November 20, 2022 at https://www.proquest.com/docview/303320065?pq-origsite=gscholar&fromopenview=true

Justice, L. M., & Ezell, H. K. (1999). Syntax and speech language pathology graduate students: Performance and perceptions. *Contemporary Issues in Communication Science and Disorders, 26*(Fall), 119–127. https://doi.org/10.1044/cicsd_26_F_119

Lee, L. L. (1974). *Developmental sentence analysis: A grammatical assessment procedure for speech and language clinicians.* Northwestern University Press.

Lee, L. L., & Canter, S. M. (1971). Developmental sentence scoring: A clinical procedure for estimating syntactic development in children's spontaneous speech. *Journal of Speech and Hearing Disorders, 36*(3), 315–340. https://doi.org/10.1044/jshd.3603.315

Long, S. (2001). About time: A comparison of computerized and manual procedures for grammatical and phonological analysis. *Clinical Linguistics & Phonetics, 15*(5), 399–426. https://doi.org/10.1080/02699200010027778

MacWhinney, B. (2008). Enriching CHILDES for morphosyntactic analysis. In C. Rowland, S. L. Fletcher, D. Freudenthal, & H. Behrens (Eds.), *Trends in corpus research: Finding structure in data.* (pp. 165–198). John Benjamins. https://doi.org/10.1075/tilar.6.09mac

MacWhinney, B. (2014). *The CHILDES project: Tools for analyzing talk.* Psychology Press.

MacWhinney, B., & Bernstein Ratner, N. (2022). Dynamic norming and open science. *Journal of Speech, Language, and Hearing Research, 65*(3), 1183–1185. https://doi.org/10.1044/2022_JSLHR-22-00019

MacWhinney, B., Roberts, J. A., Altenberg, E. P., & Hunter, M. (2020). Improving automatic IPSyn coding. *Language, Speech, and Hearing Services in Schools, 51*(4), 1187–1189. https://doi.org/10.1044/2020_LSHSS-20-00090

Miller, J. F., Andriacchi, K., & Nockerts, A. (2015). *Assessing language production using SALT software: A clinician's guide to language sample analysis* (2nd ed.). SALT Software.

Overton, C., Baron, T., Pearson, B. Z., & Ratner, N. B. (2021). Using free computer-assisted language sample analysis to evaluate and set treatment goals for children who speak African American English. *Language, Speech, and Hearing Services in Schools, 52*(1), 31–50. https://doi.org/10.1044/2020_LSHSS-19-00107

Paul, R., Norbury, C. F., & Gosse, C. (2018). *Language disorders from infancy through adolescence*. Elsevier Health Sciences.

Pavelko, S. L., & Owens, R. E. (2017). Sampling Utterances and Grammatical Analysis Revised (SUGAR): New normative values for language sample analysis measures. *Language, Speech, and Hearing Services in Schools, 48*(3), 197–215. https://doi.org/10.1044/2017_LSHSS-17-0022

Pezold, M. J., Imgrund, C. M., & Storkel, H. L. (2020). Using computer programs for language sample analysis. *Language, Speech, and Hearing Services in Schools, 51*(1), 103–114. https://doi.org/10.1044/2020_LSHSS-19-00107

Riley, G. D. (2009). *Stuttering Severity Instrument–Fourth Edition*. Pro-Ed.

Scarborough, H. S. (1990). Index of productive syntax. *Applied Psycholinguistics, 11*(1), 1–22. https://doi.org/10.1017/S0142716400008262

Templin, M. (1957). *Certain language skills in children: Their development and interrelationships [Monograph Series No. 26]*. University of Minnesota, The Institute of Child Welfare. https://doi.org/10.5749/j.ctttv2st

Yang, J. S., MacWhinney, B., & Bernstein Ratner, N. (2022). The index of productive syntax: Psychometric properties and suggested modifications. *American Journal of Speech-Language Pathology, 31*(1), 239–256. https://doi.org/10.1044/2021_AJSLP-21-00084

Yang, J. S., Rosvold, C., & Bernstein Ratner, N. (2022). Measurement of lexical diversity in children's spoken language: Computational and conceptual considerations. *Frontiers in Psychology, 13*. https://doi.org/10.3389%2Ffpsyg.2022.905789

Accessibility Content

Journal ID:	persp
ISBN:	
Journal Title:	Perspectives of the ASHA Special Interest Groups
Article ID:	2200156
Article Title:	Assessment and Therapy Goal Planning Using Free Computerized Language Analysis (CLAN) Software
Image Count:	6

Page no.	Fig/ Photo #	Thumbnail	Image Type Pick one: (D)/(R)/(S)/(I)	Image Type	Alt-text Description
	PERSP-22-00156fig1		(I)	image	A screenshot. The window in the background is titled C l a n, left square bracket, 12080, right square bracket. The window displays a short sample code for human analysis of transcripts. The window in the foreground is titled I D s and has text boxes, drop-down menus, and buttons for setting various parameters.
	PERSP-22-00156fig2		(I)	image	A screenshot of a window titled, Commands, hyphen, 31 a u g 22. The paths of the working, output, lib, and m o r lib directories are listed in the top half of the window. A drop-down box at the left center selects the option labeled P r o g s. A button at the right center is labeled Help. A large text box fills up the remaining half of the window. The buttons at the bottom left and bottom right are labeled Recall and Run. The text caption between the buttons reads: Press Up or Down Arrow Key on keyboard for Previous or Next command.
	PERSP-22-00156fig3		(I)	image	A listing of a transcript file to which grammatical and syntactic information has been added.

PERSP-22-00156fig4		(I)	image	A screenshot of 4 excel worksheets. The excel worksheets are part of a Kid Eval report.
PERSP-22-00156fig5		(I)	image	A listing of a portion of a file that analyses the developmental sentence score of a child. The full path of the file is C, colon, backslash, talk bank, backslash, clan, backslash, work. The score analysis is organized into 10 columns titled, Sentence, I P, P P, M V, S V, N G, C N J, I R, W H Q, S, and T O T, from left to right.
PERSP-22-00156fig6		(I)	image	A table listing the codes corresponding to the various categories of the English parts of speech. The categories and the codes are as follows. Adjective: a d j. Adjective-predicative: a d j, colon, p r e d. Adverb: a d v. Adverb-temporal: a d v, colon, t e m. Communicator: c o. Complementizer: c o m p. Conjunction: c o n j. Coordinator: c o o r d. Determiner Article: d e t, colon, a r t. Determiner-demonstrative: d e t, colon, d e m. Determiner-interrogative: d e t, colon, int. Determiner-numeral: d e t, colon, n u m. Determiner-possessive: d e t, colon, p o s s. Filler: f i l. Infinitive: inf. Negative: n e g. Noun: n. Noun-letter: n, colon, l e t. Noun-plurale tantum: n, colon, p t. Proper noun: n, colon, p r o p. Onomatopoeia: o n. Particle: part. Postmodifier: post. Preposition: prep. Pronoun-demonstrative: pro, colon, d e m. Pronoun-existential: pro, colon, exist. Pronoun-indefinite: pro, colon, indef. Pronoun-interrogative: pro, colon, int. Pronoun-object: p r o, colon, o b j. Pronoun-personal: pro, colon, p e r. Pronoun-possessive: pro, colon, p o s s. Pronoun-reflexive: pro, colon, r e f l. Pronoun-subject: pro, colon, s u b. Quantifier: q n. Verb: v. Verb-auxiliary: a u x. Verb-copula:

					cop. Verb-modal: mod.

Image Type Pick one: (D) Decorative, (R) Redundant, (S) Supplemental, (I) Informational