Supplemental Material S2. Results on matched utterance sets

CLAN, SALT, and SUGAR have different rules for which utterances are included in analyses. In CLAN, utterances that have unintelligible words or segments are not included in analyses, while abandoned or interrupted utterances are included. In SALT, neither utterances with unintelligible material nor abandoned or interrupted utterances are included. In SUGAR, utterances with fewer than three unintelligible words are included, as are abandoned or interrupted utterances. (Note that no program includes mazes, retraces, or repetitions, although specific coding differs across programs.) The difference in utterance inclusion rules across programs meant that we were comparing the conversation language samples from our two preschool cases across different utterance sets for each program.

In this supplement, we report analyses from CLAN, SALT, and SUGAR for the same 50-utterance set. Arbitrarily, we chose to use SALT conventions for utterance inclusion (excluding utterances that are abandoned, interrupted, or contain unintelligible segments). First, we created a copy of the SALT transcript that contained only the 50 utterances included for analysis by removing all words not analyzed by the program (mazes, retraces, repetitions, and utterances that are abandoned, interrupted, or contain unintelligible segments). We converted this SALT transcript to CLAN format using CLAN's SALTIN command, which automatically changes SALT coded transcripts to match CLAN conventions. We used the check feature and +xb command in CLAN to verify that the transcript had no errors or unknown words, in addition to proofreading the file. Finally, we copied and pasted the SALT transcription into a Microsoft Word file and removed the morpheme coding and speaker codes. We then repeated hand-coding of morphemes in the sample according to SUGAR's guidelines.

We repeated our analyses on our newly created matched transcripts. Results are reported in Table S2. In CLAN, we ran the MOR command and the KIDEVAL command to get mean length of utterance in morphemes (MLU), number of different words (NDW, called "types" in the KIDEVAL output), and total number of words (TNW, called "tokens" in the KIDEVAL output). KIDEVAL also reported the standard deviation difference of each child's values from the CHILDES database mean. In SALT, we computed standard measures with database comparisons to children within a one-year age range and from samples with the same number of utterances. The SALT report also included standard deviation differences of each child's values from the relevant database samples. In SUGAR we computed TNW and MLU using the word count feature in Microsoft Word. We compared these values to those reported by Pavelko and Owens (2017) and calculated the standard deviation difference from the sample mean for same-age children. Finally, we completed hand-coding of the sample for comparison. Again, we arbitrarily chose SALT conventions for morpheme coding. We hand-counted the number of words and morphemes in each utterance, using Excel to automatically total the results.

Table S2

Analysis Results Across for Each Program for Sam and Julia on a Matched Set of Utterances

	Sam				Julia			
	CLAN[a]	SALT[b]	SUGAR[c]	Hand-coding[d]	CLAN[a]	SALT[b]	SUGAR[c]	Hand-coding[d]
MLU Raw	3.10	2.96	3.00[e]	2.96	3.94	3.92	3.92[e]	3.92
MLU *SD* from program database mean	-1.11	-1.93	-2.41		-0.36	-1.08	-1.71	
NDW Raw	68	69			77	78		
NDW *SD* from program database mean	-1.11	-1.95			-1.08	-1.45		
TNW Raw	139	131	131	131	191	186	186	186
TNW *SD* from program database mean	-0.78	-2.04	-2.46		-0.77	-0.96	-1.54	

Note. MLU = mean length of utterance in morphemes, NDW = number of different words, TNW = total number of words. [a]Compared to free play samples from 141 children age 4;6–4;11 using KIDEVAL. [b]Compared to 50-utterance play samples from 25 children age 4;3–5;3 using SALT database. [c]Compared to conversation samples from 55 children age 4;6–4;11 reported by Pavelko and Owens (2017). [d]Completed using SALT conventions for morphological coding. [e]Values for MLU SUGAR.

As reported in Table S2, raw values for MLU differed slightly across all three programs. We analyzed the transcripts by hand to locate the source of these discrepancies. For Sam, the MLU discrepancy between SALT and SUGAR resulted from two instances of "don't". SALT conventions do not count "don't" as two morphemes. SUGAR conventions do not call for counting contractions (including "don't") as two morphemes, so we marked "don't" as two morphemes in SUGAR. CLAN also counts "don't" as two morphemes. The additional discrepancy in CLAN resulted from three instances of diminutives (i.e., one instance of "piggy" and two instances of "stuffy," or a stuffed animal) which are counted as two morphemes by CLAN's automatic coding but not by SALT or SUGAR. CLAN also counted the word "farmer" as two morphemes (i.e., "farm" and the agent suffix *–er*). Neither SALT nor SUGAR mentions the handling of agent words like "farmer," so we did not code it as a separate morpheme. For Julia, the MLU discrepancy in CLAN results compared to SUGAR and SALT came from CLAN's counting of the word "telled" as two morphemes. We chose not to separate the *–ed* morpheme in this word when hand-coding with SALT and SUGAR because it is used in error. We failed to mark the word as an error using CLAN and it automatically counted it. Adding [: told] after the word will cause CLAN to treat it as a single morpheme. Although this was an error on our part, we reported it as an example of the differences across each program. Our hand-coding of MLU equaled the SALT results because we used SALT conventions for "don't," "farmer," and

diminutives. Had we used SUGAR or CLAN conventions, it would have equaled those values (with the exception of "telled" in CLAN, which resulted from a coding error on our part).

For NDW, CLAN and SALT results differed by 1–2 words. The KIDEVAL program completes all analyses using the morphological (%MOR) tier, which treats words like "it's" and "can't" as two words ("it" and copula/auxiliary "be;" "can" and "not"). We ran the FREQ command in CLAN independently of KIDEVAL to tally the number of total and unique words on the speaker tier (i.e., what we had typed). The use of this command inflated the number of types (i.e., NDW) because each inflection of a word was counted as a unique word. For example, "do" and "doing" would count as two unique words using the FREQ command in CLAN, while SALT and CLAN's KIDEVAL would count them as the same word because the *–ing* ending is coded separately. The FREQ in CLAN command reported 77 unique words for Sam and 82 unique words for Julia. The nature of NDW made it more difficult to hand-code samples to find the source of the discrepancy between CLAN's KIDEVAL and SALT. Instead, we used a free website (https://wordcounttools.com/) to find the number of unique words. The results for the raw transcript (copied and pasted from the uncoded SUGAR transcript) matched those from CLAN's FREQ command (77 for Sam, 82 for Julia). We then edited two copies of the transcript to mimic the way that CLAN's KIDEVAL and SALT would count words. For CLAN's KIDEVAL, we removed all morphological endings (including diminutives) and changed all contractions to two words (e.g., don't = do not). We eventually located several other words that needed to be changed as well. CLAN's %MOR tier codes "mhm" as "yes," "did" and "done" as "do," and "is" as "be." For SALT, we removed all morphological endings that are marked in SALT. Using this process and the word counting website, we were able to replicate the CLAN and SALT results for NDW. Again, the differences were not due to program error, but rather to differences in the way each program marks different morphemes and computes NDW.

For TNW, SALT and SUGAR results were equal for both children. CLAN results for TNW were 8 words higher for Sam and 5 words higher for Julia. Using the FREQ command in CLAN independently of KIDEVAL, the results for the number of tokens (i.e., TNW) were the same as for SALT and SUGAR (i.e., 131 for Sam and 187 for Julia). Again, the discrepancy is caused by KIDEVAL's reading of the %MOR tier, which treats contracted words as two words.

In summary, when using the same set of utterances, values for MLU, NDW, and TNW still differed slightly across programs because of the way each program counts morphemes and completes analyses. For our two preschool cases, these differences were small and not clinically meaningful.