Title: Reliability of Computerized Language Analysis (CLAN) to Evaluate Grammatical Skills of Individuals with DS or FXS

Authors: Lisa Hilliard[1], Leonard Abbeduto[2], & Lizbeth H. Finestack[1]

Introduction: In clinical practice and research projects, clinicians and researchers use a variety of assessments to examine language samples produced by clinician populations. Frequently used measures include mean length of utterance (MLU), type token ratio (TTR), Developmental Sentence Scoring (DSS; Lee, 1974), and the Index of Productive Syntax (IPSyn; Scarborough, 1990). Measures like MLU and TTR provide general estimates of syntactic complexity and semantic diversity, respectively, and are relatively simple to hand score using a transcribed language sample. Measures such as DSS and IPSyn provide more in-depth information regarding grammatical language development. Such measures are time-consuming to complete reliably. One option for clinicians and researchers is to utilize automated language analysis software to conduct DSS and IPSyn analysis, as well as to derive MLU and TTR measures. Computerized Language Analysis (CLAN; MacWhinney, 2000) is one possible option for analyzing language samples, which requires less time than hand-scoring procedures. However, very few clinicians are currently utilizing CLAN in their practice (Finestack & Satterlund, 2018) and to date, there are no known published studies that have used CLAN to derive DSS or IPSyn study measures.

There have been some reports of the reliability of CLAN's parser for identifying grammatical averages (94% accuracy) (Sagae et al., 2010) and of comparisons between human and CLAN IPSyn scoring (94% inter-agreement reliability) (Sagae, Lavie, & MacWhinney, 2005). However, the IPSyn reliability analysis was based on two samples of typically developing children (aged 2-3 years and 8-9 years). Thus, little is known regarding the reliability of CLAN analyses based on samples of individuals with language impairment, particularly those with neurodevelopmental disabilities. Thus, the aim of this study to compare CLAN data to hand-scored data to determine whether CLAN is a reliable option for analyzing the language samples of children and adolescents more efficiently.

Method: The participants in the current study included adolescents and young adults with Down syndrome (DS; *n* = 36), adolescents and young adults with fragile X syndrome (FXS; *n* = 20), and younger children with typical cognitive and language development (TD; *n* = 21). We drew this participant sample from a larger study examining the language development of adolescents and young adults with DS or FXS, which overlaps with other published studies (Abbeduto et al., 2003; Abbeduto et al., 2006; Finestack & Abbeduto, 2010; Keller-Bell & Abbeduto, 2007; Lewis et al., 2006) and the participants included in the current study are nearly identical to those included in a study by Finestack and Abbeduto (2010). Participants with DS or FXS completed both conversational and narrative language samples. The participants who were TD completed only narrative samples. Research assistants transcribed the samples using SALT conventions. Assistants then hand-scored each sample using IPSyn and DSS. The SALT transcripts were then imputed into CLAN to be analyzed using the Talk Bank's KidEval, DSS, and IPSyn analyses.

Results: The dependent measures included SALT's MLU and TTR measures, hand-scored DSS and IPSyn scores, and the same measures derived from Talk Bank's KidEval, DSS, and IPSyn reports. We then examined each groups' performance on each study measure using Cronbach alpha correlations. We also compared the language profiles of the groups using hand-scored and CLAN-scored measures to determine if there were any differences in outcomes based on the scoring procedure. Here we report preliminary results based on IPSyn, DSS, and KidEval measures derived from conversational and narrative language samples. KidEval measures including number of utterances, MLU, and TTR, were highly correlated with $\alpha = .975$ as the lowest value correlation across all groups in both conversational and narrative contexts. Full group analyses for IPSyn showed strong correlations across all measures, with the exception of the Question/Negation sub-score ($\alpha = .544$) in the conversational context. Full

group analyses for narrative DSS showed strong correlations between hand-scored and automated Indefinite Pronoun (α = .797) and Personal Pronoun (α = .767) sub-scores, but Main Verb (α = .562), Secondary Verb (α = .457) and Conjunctions (α = .395) were not strongly correlated. Similar trends were found in the DS, FXS, and TD groups when analyzed individually.

Discussion: Preliminary results suggest that the automated scores are consistent with hand-scored scores for calculating several IPSyn and DSS sub-scores, based on both conversational and narrative language. Based on these results, some automated IPSyn and DSS sub-scores may be a viable analysis approach when examining the language skills of children and adolescents with neurodevelopmental disorders. Further analyses will be conducted to examine specific relationships between hand-scored and automated data, such as the frequency of hand-scored and automated measures falling within 2 points of one another.

References:

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.

Lee, L. (1974). *Developmental sentence analysis*. Evanston, IL: Northwestern University Press.
MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk* (3rd Edition ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Sagae, K., Lavie, A., & MacWhinney, B. (2005). Automatic measurement of syntactic development in child language. Paper presented at the Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics.

Sagae, K., Davis, E., Lavie, A., MacWhinney, B., & Wintner, S. (2010). Morphosyntactic annotation of CHILDES transcripts. Journal of Child Language, 37(3), 705-729.

Scarborough, H. S. (1990). Index of productive syntax. *Applied Psycholinguistics, 11*, 1-22.

[1] University of Minnesota Twin Cities
[2] University of California Davis