Research Week 2020

Using MLU to Evaluate the Reliability of ADOS Transcription

Heather MacFarlane, Alex Salem, Jill Dolata, Grace Lawley, Eric Fombonne

OHSU

Keywords

autism, intra-annotator reliability, computational language analysis

Abstract

Transcription of language samples is labor intensive and data on intra- and inter-annotator reliability is lacking. Mean Length of Utterance in Morphemes (MLUM) can be calculated from transcripts and used to evaluate intra-annotator reliability. We developed a computational method to determine transcriber consistency across a large speech corpus.

Methods

Module 3 of the Autism Diagnostic Observation Schedule (ADOS) was administered to 137 children (IQ > 70; fluent/phrase speech), then transcribed according to modified SALT guidelines. Transcripts of Emotions, Social Difficulties and Annoyance, and Friends and Marriage conversations were split by even and odd lines and MLUM calculated separately for each half. Intra-annotator agreement was evaluated using Intraclass Correlation (ICC).

Results

The sample comprised 66 children with ASD (mean age: 11.3 years; 80% male; mean IQ: 102) and 71 controls without ASD (mean age: 11.4 years; 58% male; mean IQ: 113). Across tasks and groups, mean MLUM ranged from 6.17 to 6.37. ICC between the even and odd MLUM was 0.732 for Emotions (95% CI: 0.644-0.801), 0.593 for Social (95% CI: 0.474-0.692), and 0.744 for Friends (95% CI: 0.658-0.81), indicative of moderate to good levels of reliability. Paired-t-tests between the two MLUM halves were all nonsignificant, indicating very good within task intra-rater agreement. ICC was comparable between the two clinical groups although reliability was higher in ASD than for controls on the three tasks. Using age and IQ median splits (11.4 years and 110, respectively), we further established that age had no discernible effect on ICC across tasks; a trend was found for lower reliability on one task among subjects with higher IQ. However, all ICCs across age and IQ groups remained in the moderate (>.50) range.

Conclusions

Calculating MLUM for two random halves within each activity provides an efficient and valid measure of intra-annotator reliability.